

**Multi-omics integration in Alzheimer's disease:
Linking genotype to phenotype through glial
activation, synaptopathy and cognitive decline**

I M P E R I A L

Eléonore Marguerite Antoinette Schneegans

Submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy

Department of Brain Sciences,
Imperial College London
UK Dementia Research Institute

January 2025

Statement of Originality

I hereby declare that, except where explicit reference is made to the work of others, all work presented in this thesis has been carried out and written by me. Where information is derived from external sources or work conducted in collaboration, it is clearly acknowledged in the declaration sections of the respective chapters. For methods involving wet lab experimentation that generated datasets, I have explicitly acknowledged the individuals responsible for performing the work. This thesis contains less than 100,000 words including appendices, bibliography, tables and equations and has less than 150 figures.

Eléonore Marguerite Antoinette Schneegans

January 2025

Copyright Declaration

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC). Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose. When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Acknowledgements

I express my gratitude to the donors and their families, whose generous contribution of human brain tissue made this research possible. I also thank the UK Brain Bank staff for making such resource available; without their support, research of this kind would not exist.

I am grateful for the financial support I received through the James and Gloria Borley Studentship Fund from Imperial College London. This work was also supported by the UK Dementia Research Institute, funded by the UK Medical Research Council, Alzheimer's Society, and Alzheimer's Research UK.

I have been mentored by inspiring scientists throughout this journey. I thank my primary supervisor, Dr Johanna Jackson, for her guidance, support and patience. Her supervision has shaped my research at every level, from developing new methods to data interpretation and communicating the results. I will be forever grateful for her strong collaborative spirit, creativity and for making our work together feel effortless. I am also deeply thankful to Professor Paul Matthews for welcoming me into his group, continuously encouraging me to consider the broader implications and translational aspects of my research.

I extend special thanks to my co-supervisor, Dr Nurun Fancy, a phenomenal bioinformatician whose expertise was essential in helping me navigate complex data analyses. Her mentorship, which began during my master's studies and continued over the past four years, has been fundamental in honing my computational skills.

I thank the entire wet lab team, particularly Emily Adair, Vicky Chau, Dorcas Cheung, and Marianna Papageorgopoulou for generating high-quality data for the Multi-omics Atlas Project, which now represents an invaluable public resource for advancing our understanding of Alzheimer's disease. I also thank Diana P. Benitez for her critical support in human tissue management.

To my family and friends, your belief in my abilities motivated me throughout my academic journey. Without a doubt, I would not have achieved this without your support. Finally, a warm thank you to my partner Gaspard Boscher, for his encouragement and love, which have carried me through every challenge of this journey.

Abstract

Alzheimer's disease (AD) is not a linear process driven solely by amyloid-beta and tau deposition but rather a dynamic, multifaceted disease involving interconnected biochemical, cellular, and clinical stages that evolve over many years and are preceded by a long prodromal phase. In this thesis, I explored how glial activation, specifically of microglia and astrocytes, contributes to AD pathogenesis by adopting a bottom-up approach from genotype to phenotype. I hypothesised that glial activation contributes to cognitive decline through its effect on synaptopathy, forming a glia-cognition link that underlies clinical heterogeneity in AD.

I integrated bulk transcriptomics and proteomics with snRNA-seq, synaptic, and spatial omics data across multiple AD cohorts. I developed and applied an R package, Omix, to integrate transcriptomic and proteomic datasets, enabling the identification of glial-specific disease modules. These modules were mapped to genotype information and validated against snRNA-seq subpopulations and imaging mass cytometry data.

Results highlighted a role for early microglial activation in the genesis of AD and later astrocyte reactivity as a determinant of clinical progression. Microglial activation preceded severe neuropathology, orchestrating local inflammation and complement-driven synaptic pruning. A neurotoxic astrocyte response was associated with accelerated cognitive decline. Astrocyte-associated synaptopathy, involving peri-synaptic astrocyte processes (PAPs), emerged as a potential driver of synapse loss. Imaging mass cytometry localised neurotoxic astrocytes preferentially around A β plaques.

This thesis provides evidence that glial activation actively shapes AD pathogenesis and suggests that targeting glial cells, particularly astrocytes, may slow cognitive decline. I nominated fluid-based candidate biomarkers for detecting individuals at risk of rapid decline, offering potential for patient stratification in clinical trials. Ultimately, this work highlights how integrative multi-omics can reveal key mechanisms, therapeutic targets, and biomarkers in complex neurodegenerative diseases.

Contents

STATEMENT OF ORIGINALITY	II
COPYRIGHT DECLARATION.....	III
ACKNOWLEDGEMENTS	IV
ABSTRACT	V
CONTENTS	VI
LIST OF TABLES.....	XII
LIST OF SUPPLEMENTARY FIGURES	XIII
ABBREVIATIONS.....	XIV
CHAPTER 1 GENERAL INTRODUCTION AND BACKGROUND	17
1.1 ALZHEIMER’S DISEASE: GENOTYPE TO PHENOTYPE	18
1.1.1 <i>Historical perspectives and overview of AD</i>	18
1.1.2 <i>AD risk factors</i>	19
1.1.3 <i>The three phases of AD</i>	21
1.2 HIGH-DIMENSIONAL OMICS DATA.....	28
1.2.1 <i>Genomics</i>	29
1.2.2 <i>Transcriptomics</i>	29
1.2.3 <i>Proteomics</i>	31
1.2.4 <i>Epigenetics</i>	32
1.3 INTEGRATIVE ANALYSES	33
1.3.1 <i>What is multi-omics integration?</i>	33
1.3.2 <i>Multi-omic integrative analyses in human AD</i>	33
1.4 THESIS OVERVIEW	40
1.5 PERMISSIONS INFORMATION CHAPTER 1	43
CHAPTER 2 OMIX: A TRANSCRIPTOMICS-PROTEOMICS INTEGRATION PIPELINE	45
ABSTRACT	46
GRAPHICAL ABSTRACT	47
2.1 INTRODUCTION.....	48
2.1.1 <i>From single platform to integrative multi-omics analyses</i>	48
2.1.2 <i>Transcriptomics-proteomics integration</i>	50
2.1.3 <i>Integration challenges</i>	51
2.1.4 <i>Aim and objectives</i>	52
2.2 OMIX PIPELINE IMPLEMENTATION	52
2.2.1 <i>Package Overview</i>	52
2.2.2 <i>Multimodal container</i>	53
2.2.3 <i>Processing and quality control</i>	53
2.2.4 <i>Single omic models</i>	55
2.2.5 <i>Multi-omics vertical integration</i>	57
2.2.6 <i>Joint Transcriptomics-Proteomics post-integration downstream analyses</i>	60
2.2.7 <i>Visualisations and reports</i>	63
2.2.8 <i>Code and documentation availability</i>	63
2.3 METHODS	64
2.3.1 <i>Publicly available AD cohort</i>	64
2.3.2 <i>Evaluation and comparison of transcriptomics-proteomics integration in AD</i>	64
2.3.3 <i>Internal data</i>	66
2.3.4 <i>10x Multiome analysis</i>	66
2.4 RESULTS.....	67

2.4.1	<i>Partial concordance between transcriptomics and proteomics</i>	67
2.4.2	<i>Multi-omics signatures provide complementary insights into AD mechanisms</i>	70
2.4.3	<i>Bulk transcriptomics-proteomics integration uncovers microglia-specific AD regulatory mechanisms</i>	73
2.5	DISCUSSION	77
2.5.1	<i>Complementarity of transcriptomics-proteomics</i>	77
2.5.2	<i>Omix streamlines integration to characterise potential disease mechanisms</i>	78
2.5.3	<i>Transcriptomics-proteomics integration reveals microglia-specific regulatory mechanisms in AD</i>	79
2.6	CONCLUSIONS AND FUTURE WORK	80
2.7	SUPPLEMENTARY TABLES	81
2.8	SUPPLEMENTARY FIGURES	81
CHAPTER 3 INTEGRATIVE MULTI-OMICS ANALYSIS OF GLIAL SIGNATURES ASSOCIATED WITH ACCELERATED COGNITIVE DECLINE IN ALZHEIMER'S DISEASE		92
ABSTRACT		93
GRAPHICAL ABSTRACT		94
3.1	INTRODUCTION	95
3.1.1	<i>Moving beyond amyloid and tau: integrating cognitive data with multi-omics in AD</i>	95
3.1.2	<i>The roles of microglia and astrocytes in AD cellular phase</i>	95
3.1.3	<i>Aim and objectives</i>	96
3.2	METHODS	97
3.2.1	<i>Brain tissue</i>	97
3.2.2	<i>Bulk transcriptomics</i>	97
3.2.3	<i>Bulk proteomics and pre-processing</i>	98
3.2.4	<i>Single nuclei transcriptomics</i>	100
3.2.5	<i>Immunohistochemistry (IHC)</i>	101
3.2.6	<i>Processing of bulk transcriptomics-proteomics data</i>	102
3.2.7	<i>Bulk multi-omics integration</i>	102
3.2.8	<i>Cross-sectional pseudotime trajectory</i>	103
3.2.9	<i>Transcriptomics-proteomics co-expression modules</i>	103
3.2.10	<i>Differential module analysis</i>	104
3.2.11	<i>Module cell type enrichment</i>	104
3.2.12	<i>Module functional enrichment</i>	104
3.2.13	<i>Projecting bulk modules and gene sets in single nuclei data</i>	105
3.2.14	<i>Microglia and astrocyte cluster differential expression analysis</i>	105
3.2.15	<i>AD GWAS enrichment analysis</i>	106
3.2.16	<i>Transcription factors target gene enrichment analysis</i>	106
3.2.17	<i>Microglia and astrocyte module specific PRS</i>	107
3.2.18	<i>Rate of cognitive decline</i>	107
3.2.19	<i>Multi-omics subtyping of early AD</i>	107
3.2.20	<i>Differential molecular subtype analysis and pathway enrichment</i>	108
3.2.21	<i>Kaplan-Meier analysis</i>	109
3.2.22	<i>Random forest model to predict pseudotime</i>	109
3.2.23	<i>SNARE proteomics</i>	110
3.2.24	<i>Linear mixed-effects model with interaction terms</i>	110
3.2.25	<i>CSF and plasma proteomics modules overlap</i>	111
3.2.26	<i>Cognitive resilience classification</i>	112
3.2.27	<i>Data availability</i>	114
3.2.28	<i>Code availability</i>	115
3.3	RESULTS	115
3.3.1	<i>Bulk transcriptomics-proteomics integration</i>	115
3.3.2	<i>Cross-sectional pseudotime analysis</i>	118
3.3.3	<i>Transcriptomics-proteomics co-expression networks show AD-correlated glial activation mechanisms and neuronal injury response</i>	119
3.3.4	<i>Microglia and astrocyte activation mechanisms are cell state-specific and enriched in AD genetic risk contributed either by transcriptomics or proteomics</i>	121

3.3.5	<i>AD polygenic risk modulates glial activation and its effects on the rate of cognitive decline</i>	125
3.3.6	<i>Glial activation is linked to synapse dysfunction and a greater rate of cognitive decline in early AD</i>	128
3.3.7	<i>Biomarkers in plasma and CSF: Identifying high-risk groups for rapid cognitive decline</i>	132
3.3.8	<i>Linking the rate of cognitive decline to the notion of cognitive resilience.....</i>	134
3.4	DISCUSSION.....	140
3.4.1	<i>Glial activation dynamics</i>	141
3.4.2	<i>Differential cognitive decline groups and their molecular correlates in early AD.....</i>	142
3.4.3	<i>Identification of potential CSF and plasma biomarkers for rapid cognitive decline</i>	143
3.4.4	<i>Study limitations</i>	144
3.5	CONCLUSIONS AND FUTURE WORK.....	145
3.6	SUPPLEMENTARY FIGURES	146
3.7	SUPPLEMENTARY TABLES.....	153
3.8	SUPPLEMENTARY FILES	154
CHAPTER 4 A SYNAPTIC-ASTROCYTIC PROTEOMIC SIGNATURE ASSOCIATED WITH SYNAPTOPATHY IN ALZHEIMER'S DISEASE		156
ABSTRACT		157
GRAPHICAL ABSTRACT		158
4.1	INTRODUCTION.....	159
4.1.1	<i>Studying synaptopathy in AD using proteomics.....</i>	159
4.1.2	<i>Reactive astrocytes, PAPs and synaptopathy in AD</i>	160
4.1.3	<i>Aim and objectives</i>	162
4.2	METHODS	163
4.2.1	<i>Synaptic and cytosolic protein extraction</i>	163
4.2.2	<i>Label-free quantitative proteomics</i>	164
4.2.3	<i>Synaptome mapping</i>	165
4.2.4	<i>Processing of proteomics data.....</i>	166
4.2.5	<i>Differential protein abundance.....</i>	166
4.2.6	<i>Functional enrichment.....</i>	167
4.2.7	<i>SynGO enrichment.....</i>	167
4.2.8	<i>Synaptic and cytosolic fractions integration.....</i>	167
4.2.9	<i>Pseudotime derivation</i>	168
4.2.10	<i>Synaptic-cytosolic proteomics co-expression modules</i>	168
4.2.11	<i>Targeted exclusion cell type enrichment.....</i>	169
4.2.12	<i>Gene set enrichment.....</i>	169
4.2.13	<i>ASC/PAP activity</i>	170
4.2.14	<i>Cell-cell communication analysis</i>	170
4.2.15	<i>Defining the GluNI⁺ interactome</i>	171
4.2.16	<i>Synaptic proteomics genetic enrichment.....</i>	171
4.2.17	<i>Data availability</i>	172
4.2.18	<i>Code availability.....</i>	172
4.3	RESULTS.....	172
4.3.1	<i>Selective vulnerability of synaptic components at different spatiotemporal stages of AD.....</i>	172
4.3.2	<i>Differential excitatory synapse subtype vulnerability in AD</i>	178
4.3.3	<i>Synaptic-cytosolic proteomics integration identifies a divergent glia-synapse relationship</i>	179
4.3.4	<i>Astrocyte signature at the synapse versus general reactivity profiles in AD</i>	189
4.3.5	<i>Linking astrocyte reactivity to synaptic resilience.....</i>	191
4.4	DISCUSSION.....	192
4.4.1	<i>Synaptic proteomics spatiotemporal changes.....</i>	193
4.4.2	<i>The impact of PAPs in AD</i>	194
4.4.3	<i>Synapse loss, synapse resilience and the role of PAPs</i>	195
4.4.4	<i>Study limitations</i>	196
4.5	CONCLUSIONS AND FUTURE WORK.....	198
4.6	SUPPLEMENTARY FIGURES	200
4.7	SUPPLEMENTARY TABLES.....	207

CHAPTER 5	SPATIAL OMIC CHARACTERISATION OF GLIAL ACTIVATION IN ALZHEIMER'S DISEASE USING IMAGING MASS CYTOMETRY	208
ABSTRACT		209
GRAPHICAL ABSTRACT		210
5.1	INTRODUCTION.....	211
5.1.1	<i>Spatial omics: integrating molecular profiling into a spatial context</i>	<i>211</i>
5.1.2	<i>Studying AD from a spatial angle.....</i>	<i>212</i>
5.1.3	<i>The plaque niche: AD's neuropathological microenvironment</i>	<i>213</i>
5.1.4	<i>Aim and objectives</i>	<i>214</i>
5.2	METHODS.....	215
5.2.1	<i>Imaging Mass Cytometry.....</i>	<i>215</i>
5.2.2	<i>Steinbock workflow & cell segmentation.....</i>	<i>216</i>
5.2.3	<i>Plaque segmentation.....</i>	<i>217</i>
5.2.4	<i>Cell distances.....</i>	<i>217</i>
5.2.5	<i>IMC data processing.....</i>	<i>217</i>
5.2.6	<i>Cell phenotyping.....</i>	<i>218</i>
5.2.7	<i>Correspondence IMC-snRNA</i>	<i>219</i>
5.2.8	<i>Cell type proportion tests via permutations</i>	<i>220</i>
5.2.9	<i>Differential protein expression analysis relative to distance from plaque distance</i>	<i>220</i>
5.2.10	<i>Density of cell types per plaque niches</i>	<i>221</i>
5.2.11	<i>Pseudotime of plaque niches.....</i>	<i>221</i>
5.2.12	<i>Cell-cell interaction</i>	<i>222</i>
5.2.13	<i>Cell densities and cell-interactions along pseudotime.....</i>	<i>222</i>
5.2.14	<i>Code availability.....</i>	<i>223</i>
5.3	RESULTS.....	223
5.3.1	<i>IMC microglia and astrocyte subpopulations recapitulate snRNA states.....</i>	<i>223</i>
5.3.2	<i>Molecular and cellular footprinting of plaque niches in AD</i>	<i>227</i>
5.3.3	<i>Glial activation states show spatial variability relative to plaque niches.....</i>	<i>229</i>
5.3.4	<i>Imbalance in astrocyte states and altered cell-cell interactions within the plaque microenvironment along the plaque trajectory.....</i>	<i>231</i>
5.4	DISCUSSION.....	235
5.4.1	<i>Distinct glial subpopulation and spatial alterations in AD plaque microenvironment</i>	<i>236</i>
5.4.2	<i>Pseudotemporal dynamics reveal astrocyte-microglia shifts towards neurotoxicity in plaque niches</i>	<i>237</i>
5.4.3	<i>Study limitations</i>	<i>238</i>
5.5	CONCLUSIONS AND FUTURE WORK.....	239
5.6	SUPPLEMENTARY FIGURES	240
CHAPTER 6	CONCLUDING REMARKS AND FUTURE WORK.....	242
6.1	MULTI-OMICS INTEGRATION IN AD: KEY RESEARCH OUTCOMES.....	243
6.1.1	<i>Integration scenarios encountered in this thesis</i>	<i>243</i>
6.1.2	<i>Molecular characterisation of glial activation in AD.....</i>	<i>246</i>
6.1.3	<i>Bridging molecular mechanisms and phenotype through resilience and vulnerability</i>	<i>251</i>
6.2	OVERALL CHALLENGES	251
6.2.1	<i>Bulk, single nuclei, spatial, temporal: reflecting on multi-level resolutions challenges.....</i>	<i>252</i>
6.2.2	<i>Chain of inferences and causality.....</i>	<i>253</i>
6.3	TRANSLATIONAL APPLICATIONS.....	254
6.3.1	<i>Patient stratification and monitoring.....</i>	<i>254</i>
6.3.2	<i>Novel therapies</i>	<i>256</i>
6.4	CLOSING PARAGRAPH.....	258
6.5	SUPPLEMENTARY FIGURES	259
BIBLIOGRAPHY		260

List of Figures

<i>Figure 1-1 The genetic landscape of Alzheimer's disease.....</i>	<i>21</i>
<i>Figure 1-2 Timing of major AD pathophysiological events in relation to clinical course.....</i>	<i>22</i>
<i>Figure 1-3 High-dimensional multi-omics profiling of post mortem samples in AD</i>	<i>28</i>
<i>Figure 1-4 RNA-seq workflows.....</i>	<i>30</i>
<i>Figure 1-5 Proteomics workflow</i>	<i>32</i>
<i>Figure 1-6 Overview of thesis: Multi-omics integration in AD: Linking genotype to phenotype through glial activation, synaptic dysfunction and cognitive decline.....</i>	<i>40</i>
<i>Figure 2-1 The Omix pipeline.....</i>	<i>47</i>
<i>Figure 2-2 The landscape of vertical integrative models</i>	<i>50</i>
<i>Figure 2-3 Schematic view of MOFA model.....</i>	<i>58</i>
<i>Figure 2-4 Schematic view of iCluster model.....</i>	<i>59</i>
<i>Figure 2-5 Schematic view of the DIABLO model.....</i>	<i>60</i>
<i>Figure 2-6 Differential analyses reveal partial concordance between transcriptomics and proteomics.....</i>	<i>69</i>
<i>Figure 2-7 Multi-omics approaches partially overlap with single platform signatures and point to novel markers and pathways</i>	<i>71</i>
<i>Figure 2-8 Multi-omics approaches are biologically relevant, predictive and robust.....</i>	<i>73</i>
<i>Figure 2-9 Integrated molecular profile of AD highlight complex regulatory mechanisms in microglia.....</i>	<i>77</i>
<i>Figure 3-1 A Comprehensive Multi-Regional, Multi-Omic, Multi-Resolution analysis.....</i>	<i>94</i>
<i>Figure 3-2 Optimisation of MOFA integration</i>	<i>117</i>
<i>Figure 3-3 Deriving pseudotime from latent factors</i>	<i>118</i>
<i>Figure 3-4 Transcriptomics-proteomics co-expression network define AD-correlated glial activation mechanisms and neuronal injury response.....</i>	<i>120</i>
<i>Figure 3-5 Functional characterisation of microglia and astrocyte bulk modules within their sub-cellular context.....</i>	<i>123</i>
<i>Figure 3-6 Microglia and astrocyte activation mechanisms are enriched in AD genetic risk</i>	<i>125</i>
<i>Figure 3-7 AD polygenic risk modulates glial activation directly and indirectly</i>	<i>127</i>
<i>Figure 3-8 Early AD ROSMAP cohort stratification via multi-omics clustering.....</i>	<i>128</i>
<i>Figure 3-9 Identified glial activation mechanisms are related to the trajectory of cognitive decline and synapse dysfunction.....</i>	<i>132</i>
<i>Figure 3-10 Biomarkers of glial activation and synapse dysfunction in CSF and plasma as potential biomarkers for Alzheimer's accelerated progression.....</i>	<i>133</i>
<i>Figure 3-11 Defining the cognitive resilience score.....</i>	<i>138</i>
<i>Figure 3-12 Cognitive resilience classification in models with and without neuro glia biomarkers</i>	<i>139</i>
<i>Figure 4-1 Chapter 4 Graphical abstract.....</i>	<i>158</i>

<i>Figure 4-2 Dysregulated tripartite synapse in Alzheimer's disease</i>	<i>162</i>
<i>Figure 4-3 Spatiotemporal model of Alzheimer's disease.....</i>	<i>173</i>
<i>Figure 4-4 Differential synaptic protein expression across cortical regions with AD.....</i>	<i>176</i>
<i>Figure 4-5 Up-regulated proteins at the synapse are enriched in AD genetic risk.....</i>	<i>177</i>
<i>Figure 4-6 GluN1⁺ synapse subtypes are differentially affected by AD pathology in the MTG</i>	<i>179</i>
<i>Figure 4-7 Synaptic-cytosolic proteomics integration unravels divergent glia-synapse dynamics</i>	<i>182</i>
<i>Figure 4-8 Summed S2 and S4 synapse subtype loss is associated with synaptic proteomic differences</i>	<i>184</i>
<i>Figure 4-9 Astrocyte proteome enrichment at the synapse is dominated by astrocyte junction (ASC) and peri-synaptic astrocytic process (PAP) proteins.....</i>	<i>187</i>
<i>Figure 4-10 Astrocyte to synaptic targets communication</i>	<i>188</i>
<i>Figure 4-11 Consensus reactive astrocyte signatures derived in this thesis versus the literature</i>	<i>190</i>
<i>Figure 4-12 Consensus reactive astrocyte signatures interactive network of pathways and genes</i>	<i>191</i>
<i>Figure 4-13 Astrocyte reactivity at the synapse worsens S2_S4 synaptic resilience</i>	<i>192</i>
<i>Figure 5-1 Chapter 5 Graphical abstract.....</i>	<i>210</i>
<i>Figure 5-2 Spatial omic characterisation of glial activation in AD using imaging mass cytometry.....</i>	<i>226</i>
<i>Figure 5-3 IMC allows spatial characterisation of amyloid plaques and cells</i>	<i>229</i>
<i>Figure 5-4 Microglia and Astrocyte are predominantly found in the niche.....</i>	<i>230</i>
<i>Figure 5-5 Spatial variability in protein expression.....</i>	<i>231</i>
<i>Figure 5-6 Evolution of cell-cell interactions within the plaque niche along pseudotime....</i>	<i>235</i>
<i>Figure 5-7 Neurotoxic astrocytes relate to the bulbous morphology characterised in an AD mouse model.....</i>	<i>235</i>
<i>Figure 6-1 Thesis graphical summary.....</i>	<i>242</i>

List of Tables

<i>Table 1-1 Literature review of integrative multi-omic analyses in human AD</i>	<i>39</i>
<i>Table 2-1 Processing steps implemented by Omix</i>	<i>54</i>
<i>Table 2-2 Downstream analysis modules implemented in Omix</i>	<i>62</i>
<i>Table 2-3 Comparison of Omix to other multi-omics software packages</i>	<i>63</i>
<i>Table 3-1 Antibodies used for Immunohistochemistry.....</i>	<i>101</i>
<i>Table 3-2 CSF candidates.....</i>	<i>113</i>
<i>Table 4-1 Literature derived astrocyte gene sets.....</i>	<i>170</i>
<i>Table 5-1 IMC antibody panel.....</i>	<i>216</i>
<i>Table 5-2 Markers chosen to compute IMC-snRNA correspondence scores</i>	<i>220</i>
<i>Table 5-3 Observed and expected cell proportions</i>	<i>225</i>
<i>Table 5-4 Changes in niches' microglia/astrocyte density ratio</i>	<i>233</i>

List of Supplementary Figures

<i>Supplementary figure 2-1 Single omics analyses report.....</i>	<i>82</i>
<i>Supplementary figure 2-2 Quality control report.....</i>	<i>83</i>
<i>Supplementary figure 2-3 Pathway analysis report.....</i>	<i>84</i>
<i>Supplementary figure 2-4 Multi-omics report</i>	<i>85</i>
<i>Supplementary figure 2-5 Higher relative correlation of protein modules to global pathology</i>	<i>86</i>
<i>Supplementary figure 2-6 Multi-omics signatures of AD pathology</i>	<i>88</i>
<i>Supplementary figure 2-7 MSBB module replication</i>	<i>89</i>
<i>Supplementary figure 2-8 Weighted Nearest Neighbour (WNN) UMAP visualisation of 10x multiome RNA+ATAC from internal cohort.....</i>	<i>89</i>
<i>Supplementary figure 2-9 Differentially accessible peaks.....</i>	<i>90</i>
<i>Supplementary figure 2-10 DNA accessibility in the chr19:35931069-35932252 region and TYROBP expression.....</i>	<i>90</i>
<i>Supplementary figure 3-1 Module eigenvalue validation in external cohorts (ROSMAP, MSBB).....</i>	<i>146</i>
<i>Supplementary figure 3-2 Module trajectories alongside pseudotime and association to neuropathology</i>	<i>148</i>
<i>Supplementary figure 3-3 Neuronal subpopulations and neuronal bulk module projections</i>	<i>149</i>
<i>Supplementary figure 3-4 ROSMAP cohort in early AD – multi-omics clustering</i>	<i>150</i>
<i>Supplementary figure 3-5 Random forest model to predict pseudotime.....</i>	<i>151</i>
<i>Supplementary figure 3-6 Disease modules and cognitive resilience in early AD ROSMAP.....</i>	<i>152</i>
<i>Supplementary figure 4-1 Synapse enrichment in human post-mortem cortical tissue.....</i>	<i>200</i>
<i>Supplementary figure 4-2 Synaptome mapping graphical abstract</i>	<i>201</i>
<i>Supplementary figure 4-3 Semi-supervised approach for pseudotime derivation.....</i>	<i>202</i>
<i>Supplementary figure 4-4 MOFA factor SynGO enrichments</i>	<i>203</i>
<i>Supplementary figure 4-5 Prioritised module eigenvalues regional differences stratified by case-control status</i>	<i>204</i>
<i>Supplementary figure 4-6 Inverse correlation between synaptic and astrocyte modules</i>	<i>204</i>
<i>Supplementary figure 4-7 PAP & ASC gene set activity</i>	<i>205</i>
<i>Supplementary figure 4-8 Literature reactive astrocytes signatures: 20 most common single, pair and gene trio.....</i>	<i>206</i>
<i>Supplementary figure 4-9 In vitro validation of astrocyte colocalisation with synaptoneurosomes</i>	<i>207</i>
<i>Supplementary figure 5-1 IMC spillover correction.....</i>	<i>240</i>
<i>Supplementary figure 5-2 Niche level clustering.....</i>	<i>241</i>
<i>Supplementary figure 5-3 Representative images of plaques and cells.....</i>	<i>241</i>
<i>Supplementary figure 6-1 Microglia module eigenvalue stratified by MMSE and Braak groups</i>	<i>259</i>

Abbreviations

A β : Amyloid- β
AD: Alzheimer's Disease
ADNI: Alzheimer's Disease Neuroimaging Initiative
APP: Amyloid precursor protein
ASC: Neuron-astrocyte junction proteins
ATAC-seq: Assay for Transposase-Accessible Chromatin using sequencing.
ATN: Amyloid, Tau, Neurodegeneration classification
AUC: Area Under the Curve
BH: Benjamini-Hochberg
ChIP-seq: Chromatin Immunoprecipitation Sequencing
CDR: Clinical Dementia Rating
CR: Cognitive Resilience
CRM: Cytokine and Chemokine Response
CSF: Cerebrospinal Fluid
CV: Common Variant
DAM: Disease-associated microglia
DDA: Data-Dependent Acquisition
DE: Differential Expression
DEG: Differentially expressed genes
DEP: Differentially expressed proteins
DIA: Data-Independent Acquisition
EOAD: Early-Onset Alzheimer's Disease
EWCE: Expression-Weighted Cell Type Enrichment
eQTL: Expression Quantitative Trait Loci
FDR: False Discovery Rate
FFPE: Formalin-Fixed Paraffin-Embedded
GFAP: Glial fibrillary acidic protein
GluN1: Glutamate ionotropic receptor NMDA type subunit 1
GWAS: Genome-Wide Association Studies
HLA: Human Leukocyte Antigen
IHC: Immunohistochemistry
IF: Immunofluorescence
IMC: Imaging Mass Cytometry
IRM: Interferon response microglia
LASSO: Least Absolute Shrinkage and Selection Operator
LOAD: Late-Onset Alzheimer's Disease
LPS: Lipopolysaccharides
MAF: Minor Allele Frequency
MAPT: Microtubule-associated protein tau
MCI: Mild Cognitive Impairment
ME: Module Eigenvalue
MFG-E8: Milk Fat Globule-EGF Factor 8

MMSE: Mini-Mental State Examination
MoCA: Montreal Cognitive Assessment
MOFA: Multi-Omics Factor Analysis
MRI: Magnetic Resonance Imaging
MS: Mass Spectrometry
MS/MS: Tandem Mass Spectrometry
MSBB: Mount Sinai Brain Bank
MSE: Mean Squared Error
MTG: Mid Temporal Gyrus
NDC: Non-Demented Controls
NMF: Non-Negative Matrix Factorization
NFT: Neurofibrillary Tangles
PAPs: Peri-Synaptic Astrocytic Processes
PCA: Principal Component Analysis
PET: Positron Emission Tomography
PFC: Prefrontal Cortex
PHF1: Paired Helical Filament 1
pTau: Phosphorylated Tau
PIG: Plaque-Induced Gene
PMI: Post-Mortem Interval
PRS: Polygenic Risk Score
RF: Random Forest
ROSMAP: Religious Orders Study and Memory and Aging Project
ROI: Region Of Interest
RNA-seq: RNA Sequencing
scRNA-seq: Single-cell RNA Sequencing
snRNA-seq: Single-nuclei RNA Sequencing
SNP: Single Nucleotide Polymorphism
SNARE: Soluble N-ethylmaleimide-sensitive factor Attachment Protein Receptor-mediated exocytosis
SNF: Similarity Network Fusion
sGCCA: Sparse Generalized Canonical Correlation Analysis
sMB-PLS: Sparse Multi-Block Partial Least Squares
SMOTE: Synthetic Minority Over-sampling Technique
sPLS: Sparse Partial Least Squares
SOM: Sensorimotor
STAT3: Signal Transducer and Activator of Transcription 3
SynPER: Synaptic Protein Extraction Reagent
TF: Transcription Factor
TMT: Tandem Mass Tag
TOM: Topological Overlap Matrix
TOF: Time of Flight
TREM2: Triggering Receptor Expressed on Myeloid Cells 2
UMAP: Uniform Manifold Approximation and Projection
VC: Visual Cortex
VST: Variance Stabilizing Transformation
WGCNA: Weighted Gene Correlation Network Analysis
WGS: Whole Genome Sequencing

Chapter 1 General introduction and background

1.1 Alzheimer's disease: genotype to phenotype

This section provides a historical retrospective of Alzheimer's disease (AD) from its discovery in 1906 to today. It presents the latest findings on genetic and environmental risk factors and their influence on disease risk. Additionally, it outlines a biological definition of AD, emphasising its biochemical hallmarks and cellular phase. Finally, the section examines the clinical characteristics of AD, which are not always synonymous with biologically defined AD¹.

1.1.1 Historical perspectives and overview of AD

In 1906, Alois Alzheimer started research on Auguste Deter, a 51-year-old woman suffering from memory loss and hallucinations, establishing the groundwork for what would eventually be known as AD. Initially acknowledged as “severe cognitive and behavioural disturbances” through clinical observations, it was only after her death that Dr Alzheimer discovered the presence of amyloid plaques and neurofibrillary tangles, now considered to be the hallmarks of the disease. These discoveries contributed to the growing movement of neuropathology as a scientific discipline, influenced by pioneers such as Jean-Martin Charcot and Santiago Ramón y Cajal, who advanced our understanding of the nervous system's cellular structure. Despite these early insights, research into AD stagnated for much of the 20th century.

In the 1980s, the proteins forming amyloid plaques and neurofibrillary tangles were identified as amyloid- β ($A\beta$)² and hyperphosphorylated tau³. This work laid the foundation for the amyloid cascade hypothesis⁴, which posits that the accumulation of $A\beta$ triggers a cascade of events leading to intraneuronal tau pathology, neuronal dysfunction and loss, and ultimately, cognitive decline. The 1990s brought progress in genetics, identifying mutations in the *APP*⁵, *PSEN1* and *PSEN2* genes^{6,7} that cause familial, early-onset AD (EOAD). Concurrently, the *APOE* gene was identified as a genetic risk factor for late-onset AD, with the *APOE- ϵ 4* allele substantially increasing susceptibility⁸. Concurrently, these genetic discoveries facilitated the development of animal models for studying AD mechanisms.

Over the past two decades, the scope of AD research, specifically in humans, has expanded dramatically with advances in genomics, transcriptomics, proteomics, single-cell technologies,

and imaging techniques such as positron emission tomography (PET) and magnetic resonance imaging (MRI). In the genomics category, genome-wide association studies (GWAS) have revealed over 40 genetic risk loci, predominantly implicating microglial pathways, underscoring neuroinflammation's role in the progression of AD⁹. Once considered “passive responders”, glial cells are now thought to contribute to the onset and spread of AD pathology¹⁰. Parallel advances in transcriptomics, proteomics, and single-cell technologies have revealed networks of molecular and cellular interactions and moved the research field from a neuron-centric view to a system-level understanding of the disease.

Today, despite advances in disease-modifying anti-amyloid monoclonal antibodies, AD cannot be cured. In 2019, 57 million people were estimated to suffer from dementia across the world, with AD being the most common cause. In the US, 1 in 6 individuals aged 75-84 and 1 in 3 aged 85 years or older were estimated to be affected by AD in 2020¹¹. As the population ages, its prevalence is projected to triple by 2050¹. This estimate is conservative, as a biological biomarker-based definition of AD indicates a prevalence three times higher than that based on clinical criteria¹. Imaging and fluid biomarkers are increasingly being used to detect AD pathology signs in asymptomatic individuals during the preclinical phase before cognitive symptoms arise, as well as in symptomatic individuals. These are paving the way for early detection and intervention and remain active research areas. If these early strategies are not implemented, the increasing life expectancy will result in a larger proportion of the population reaching ages at higher risk for dementia and potentially also longer periods of living with dementia¹.

1.1.2 AD risk factors

The 2024 Lancet Report for *Dementia Prevention, Intervention and Care* details 14 environmental and potentially modifiable lifestyle risk factors, including early-life (Less education), mid-life (Hearing loss >High LDL cholesterol >Depression >Traumatic brain injury >Physical inactivity >Diabetes >Smoking >Hypertension >Obesity >Excessive Alcohol) and late-life factors (Social isolation >Air pollution >Visual Loss), that altogether account for 35-45% of the lifetime risk of developing AD¹². The remaining risk, largely attributable to

genetic factors, is estimated to contribute between 55–65% of overall disease liability, though heritability estimates derived from population-based genetic studies can range as high as 58-79%, depending on the methodology used⁹. While rare (MAF<1%) autosomal dominant mutations in *APP*, *PSEN1*, and *PSEN2* genes can cause familial early-onset Alzheimer's disease (EOAD), they represent less than 5% of all AD cases (**Figure 1-1**). These mutations accelerate the production or aggregation of A β , triggering early and aggressive disease phenotypes that can manifest before the age of 65. Research on EOAD has played a crucial role in advancing knowledge of the amyloid cascade hypothesis and guiding anti-amyloid therapeutic approaches.

In contrast, the most common form of AD (95% of cases), sporadic late-onset AD, is a complex polygenic disorder influenced by the cumulative effects of multiple genetic variants and environmental factors. GWAS have identified over 80 common risk loci associated with LOAD, highlighting the roles of diverse biological pathways beyond A β metabolism, and found to largely converge on microglia response to amyloid plaque pathways⁹ (**Figure 1-1**).

The genetic architecture of LOAD is polygenic and complex, involving a spectrum of variants ranging from common SNPs, generally with small effects, to rare mutations with moderate-to-high effect sizes. The APOE- ϵ 4 remains one of the strongest known genetic risk factor—it is a common coding variant that increases risk by approximately threefold in heterozygous carriers and twelvefold in homozygous carriers¹³. APOE- ϵ 4 is associated with impaired A β clearance and increased tau-mediated neurodegeneration. Another notable example is the *TREM2* R47H variant, a rare coding mutation that increases AD risk by ~2-3 fold¹⁴ and affects microglial activation and phagocytic function. Importantly, most variants associated with AD are non-coding and lie within regulatory regions, such as promoters near genes like *BIN1* and *CLU*, or within distal enhancers that modulate gene expression through 3D chromatin interactions.

Although these variants individually confer only modest effects, their cumulative contribution can be substantial when integrated into a polygenic risk score (PRS). PRS aggregates the effects of numerous common genetic variants and has emerged as a powerful tool to stratify individuals by risk and to predict the likelihood of developing AD¹⁵. Overall, the genetic landscape of AD offers a framework for understanding how molecular and cellular mechanisms

drive disease progression, and current research efforts now integrate PRS with biomarker data to enhance disease risk prediction¹⁶.

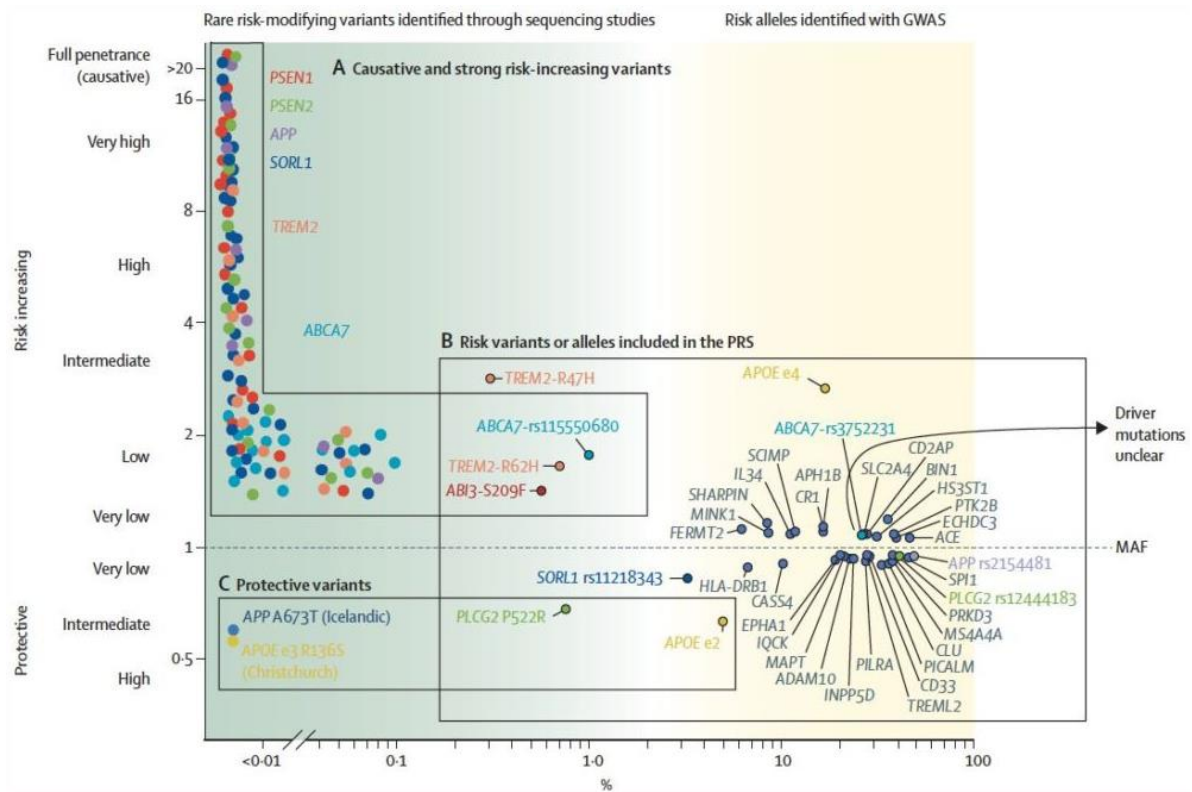


Figure 1-1 The genetic landscape of Alzheimer's disease

Image reproduced from *The Lancet*, Volume 397, Issue 10284, 1577 – 1590, Alzheimer's disease Scheltens, Philip et al. (2021)¹⁷, with permission of the rights holder Elsevier (Permissions information chapter 1). x-axis (%) MAF: Minor Allele Frequency

1.1.3 The three phases of AD

De Strooper et al. propose a framework for AD progression driven by complex biochemical cellular and clinical processes across three phases¹⁸. Each phase constitutes inflexion points yet remains dynamic, with feedback and feedforward loops that allow for bidirectional interactions between phases, ultimately shaping the disease trajectory. These include (1) the biochemical phase, characterised by A β and tau pathologies, and (2) the cellular phase, marked by glial activation, lipid dysregulation, and synaptic dysfunction¹⁸. The cellular phase appears central in determining the onset of clinical symptoms, as the failure of glial cells to maintain homeostasis may drive the transition from cellular dysfunction to cognitive decline that defines the (3) clinical phase^{19,20}. Understanding this complex cellular response that drives

the prodromal phase of AD, with precise temporal and spatial resolution, is necessary to develop a comprehensive framework that captures AD as a disorder of physiologically “disrupted cellular connectivity”¹⁸ rather than isolated molecular dysfunction. **Figure 1-2** summarises the timing of the major AD pathophysiological events from the biochemical and cellular phases in relation to the clinical course of AD.

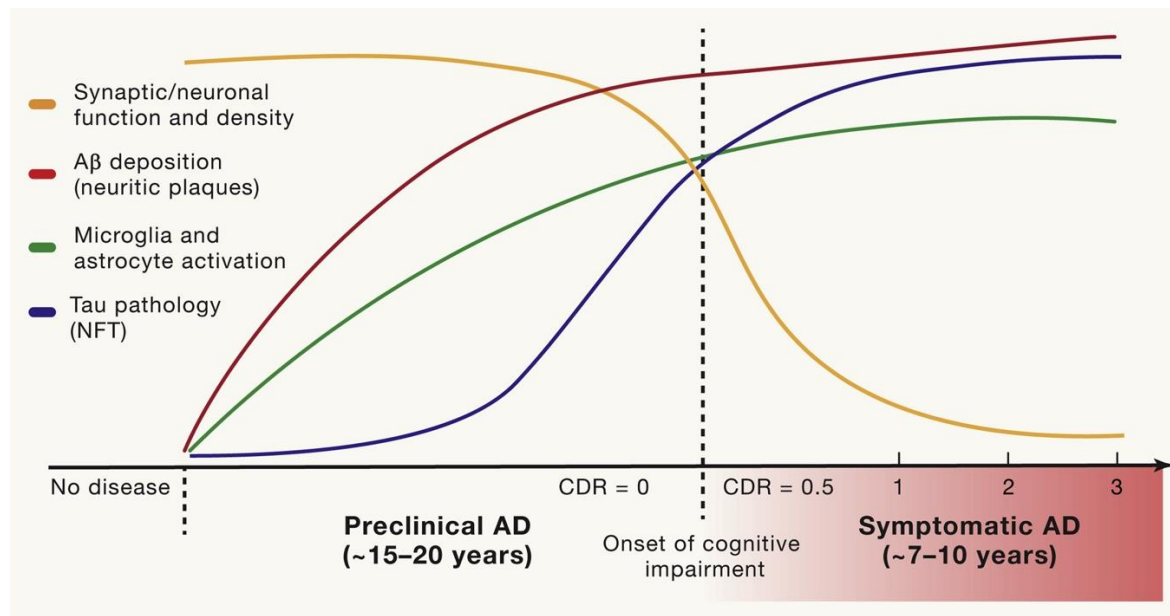


Figure 1-2 Timing of major AD pathophysiological events in relation to clinical course
 Image reproduced from Cell, Volume 179, Issue 3, 312 – 339, Alzheimer Disease: An Update on Pathobiology and Treatment Strategies, Long & Holtzman (2019)²¹, with permission of the rights holder, Elsevier (**Permissions information chapter 1**). Astrocytic activation precedes tau pathology but follows Aβ deposition.

1.1.3.1 The biochemical phase

The biochemical phase of AD is characterised by the accumulation of Aβ peptides, particularly Aβ₄₂, which are derived from the amyloid precursor protein (APP) and cleavage by β- and γ-secretase enzymes. The Aβ molecules tend to misfold and aggregate, initially forming soluble oligomers and eventually insoluble fibrils, which deposit in the brain as extracellular amyloid plaques. In brain sections, Aβ plaques are detectable via immunostaining via anti-Aβ markers like 4G8²². Genetic mutations in *APP*, *PSEN1*, and *PSEN2* alter the proteolytic processing of APP, leading to increased production of aggregation-prone Aβ peptides and promoting amyloid accumulation. The amyloid hypothesis proposes that Aβ deposition causally initiates a pathogenic, toxic cascade responsible for neurodegeneration in AD⁴. Critiques of the theory have been summarised elsewhere²³ but include (1) the lack of correlation between amyloid

plaque and clinical severity, as A β can be found in many asymptomatic individuals; or (2) repeated failures of anti-A β therapies, (3) multifactorial mechanisms including tau, neuroinflammation and metabolic dysfunction. Another key feature of the biochemical phase is the abnormal hyperphosphorylation of the tau protein, encoded by the *MAPT* gene. In healthy brains, tau stabilises microtubules in neurons. However in AD, hyperphosphorylated tau accumulates and forms neurofibrillary tangles (NFT) inside neurons, disrupting cell function and leading to cell death. The Braak staging system describes the spatiotemporal progression of tau pathology, beginning in the entorhinal cortex and advancing through the hippocampus and neocortex, correlating strongly with clinical symptoms²⁴. Phosphorylated tau (pTau) neuropathology can be detected in brain sections using specific markers such as AT8²⁵ and PHF1²⁶, which identify different phosphorylated tau species, Ser202/Thr205/Ser208 and Ser396/Ser404, respectively.

1.1.3.2 The cellular phase

The biochemical phase's proteopathic stress leads to the cellular phase of AD¹⁸. In this phase, we start to see the activation of glial cells, including microglia, astrocytes, oligodendrocytes and endothelial cells. Among these, microglia and astrocytes play a central role in driving the cellular phase and are the cells of focus of this thesis. Microglia, the brain's resident immune cells from a unique lineage, are highly motile and continuously monitor the brain parenchyma to detect changes. In contrast, astrocytes, the brain's most abundant glial cells, develop from a common progenitor with neurons and oligodendrocytes. They tile the brain and extend their processes near synapses to form the "tripartite synapse", where they play a crucial role in synapse formation, maturation and maintaining neuronal health²⁷. Both cell types are critical for normal brain development and respond rapidly to disease, infection, and trauma. Reactive gliosis, an umbrella term designating the activation of microglia and astrocytes, is known to support homeostasis initially but later to worsen neurodegeneration. Indeed, the initial inflammatory response contributes to sustaining homeostasis but then turns into a chronic neuroinflammation state characterised by the failure of clearance mechanisms, lipid dyshomeostasis and synaptic dysfunction, altogether exacerbating neuronal damage and creating a vicious cycle of neurodegeneration²⁰. Finally, the cellular phase is also marked by

important vascular changes²⁸, mostly involving the breakdown of the blood-brain barrier, though these are not the focus of this thesis.

1.1.3.2.1 Microglia activation

Microglia activation is triggered by the accumulation of A β in the brain and plays a crucial role in the clearance of A β and maintaining brain homeostasis. Still, the human microglial response to A β pathology is complex, as exemplified by the work of Mancuso et al.²⁹, which describes the multitude of microglial states, ranging from disease-activated microglia (DAM) to antigen response (HLA) to interferon response (IRM). A distinctive pro-inflammatory cytokine and chemokine (CRM) response is mounted against oligomeric A β , known to appear earlier than amyloid plaques. The Triggering Receptor Expressed on Myeloid Cells 2 (TREM2) plays a crucial role in microglial activation and has a protective effect in AD. Microglia with the *TREM2*^{R47H} variant (loss of function variant) exhibit reduced synaptic density and upregulation of several complement cascade components, indicating that dysfunctional synaptic pruning may be one of the underlying mechanisms³⁰. Another study by Fancy et al. suggested that impaired amyloid clearance by microglia with the *TREM2*^{R47H} leads to pathological astrocyte activation³¹. Microglia activation can be detrimental, contributing to synaptic loss, altering neuronal network function through changes in TNF- α , C1q and Il-1 α , and causing early neuronal dysfunction³². Conversely, *APOE*- ϵ 4 or the *TREM2*^{R47H} variant have also been found to hamper the HLA profile, suggesting a protective role of this response against AD²⁹. Still, the expert consensus defined in the paper *Defining microglial states and nomenclature: a roadmap to 2030*³³ stresses the need to move away from the “good versus bad” view of microglial states in disease and that these should be studied from a multi-dimensional angle instead. Microglial states like DAM are complex and plastic and stem from a combination of the cells’ epigenome, transcriptome, and proteome in a given pathological context, resulting in a specific functional role.

1.1.3.2.2 Astrocyte reactivity

In response to disease or injury, astrocytes undergo reactive astrogliosis, marked by heightened expression of intermediate filament proteins (i.e. glial fibrillary acidic protein [GFAP],

Vimentin [VIM]) and immune response genes, accompanied by morphological and functional remodelling (hypertrophic so-called “bushy” astrocytes)^{34,35}. In acute models of AD, at least two types of reactive astrocytes were formally identified: a neurotoxic type known as A1, triggered acutely by lipopolysaccharides (LPS)-activated inflammatory microglia, and a neuroprotective type called A2, which is induced acutely by ischemia as initially proposed by Liddelow et al. in 2017³². These disease-activated astrocytes (DAA) induced by pro-inflammatory-activated microglia are thought to exacerbate oxidative stress and contribute to neuronal death, mainly through the release of neurotoxic factors, including pro-inflammatory cytokines, chemokines and NF-κB-activated astrocyte release of complement C3 and C4^{36,37}. At very advanced stages of AD, this sustained pro-inflammatory astrocyte reactivity may also lead to a dystrophic phenotype³⁸. The concurrent dysfunction in normal supportive functions, namely calcium homeostasis and glutamate transport, have been linked to excitotoxicity, synapse loss and cognitive decline²⁷. Nevertheless, one should remember that in true disease conditions, the spectrum of astrocyte reactivity is much more granular than a dichotomic detrimental-protective A1-A2 response. This complexity was first evident from astrocytes displaying a mixed or “pan-reactive” profile, with overlapping transcriptional characteristics of both A1 and A2 types in the study that initially proposed the A1/A2 framework³². Then, in a 2022 study, Aβ and Tau were shown to induce distinct, partly overlapping astrocyte phenotypes with both harmful and adaptive-protective signals³⁹. Recently, Serrano-Pozzo et al. demonstrated in a landmark snRNA-seq study of human AD the presence of a spectrum of astrocyte subpopulations spanning homeostatic, intermediate, reactive and a newly discovered “exhausted” phenotype⁴⁰. This reinforced the prototypical homeostatic-to-reactive transition trajectory, yet one that exists alongside a distinct trajectory that can also lead to exhausted phenotypes. These underpin a complex transcriptomic response from adaptive to maladaptive/dysfunctional proposed to be exacerbated by chronic exposure to Aβ, pTau, neurodegeneration, as well as through non-cell-autonomous processes of glial activation.

1.1.3.2.3 Microglia-astrocyte crosstalk

A growing body of research is now investigating the crosstalk between microglia and astrocytes and how this crosstalk itself contributes to synaptopathy and neuronal loss in AD^{10,41}. In addition to the neuroinflammatory dynamics highlighted previously, the glia crosstalk governs key pathological processes, including the degradation of Aβ aggregates, where microglia-

astrocyte interactions facilitate aggregate clearance⁴². A recent spatial transcriptomics study by Mallach et al. suggests that microglial accumulation around amyloid plaques induces astrocytes to adopt neurotoxic phenotypes, disrupting neuronal signalling, reducing glutamatergic signalling, and promoting synaptic dysfunction⁴³. Moreover, reciprocal signalling between microglia and astrocytes activates the complement system, which is known to be involved in synapse pruning⁴². Studies like those of Tzioras et al. have shown increased synaptic protein engulfment by glial cells in AD, further aggravated by proximity to amyloid plaques and APOE-ε4 risk alleles⁴⁴. Additionally, the phagocytic mediator milk fat globule-EGF factor 8 (MFG-E8), secreted by astrocytes, has been identified as a critical regulator of microglial synapse engulfment⁴⁵. Its therapeutic targeting has shown potential in mitigating synapse loss in amyloidosis models⁴⁴. Recent findings comparing AD and asymptomatic AD patients with equivalent neuropathological loads of Aβ plaques and tangles have highlighted striking differences in glial activation^{19,46–48}, providing evidence that reactive gliosis can no longer be considered a secondary event caused by neurodegeneration, but rather as an active player in AD pathogenesis¹⁰. Differential glial activation may also partly explain some of the heterogeneity in clinical outcomes, including variations in the rate of disease progression, as recently suggested⁴⁹.

1.1.3.2.4 Synaptopathy and neuronal loss

Finally, the cellular phase of AD is marked by complex and dynamic neuronal dysfunction, including excitatory and inhibitory network imbalances, synaptic remodelling, and compensatory changes. Synaptopathy is the umbrella term designating a state associated with the dysfunction, structural alteration, or loss of synapses, which impairs neuronal communication and contributes to the progression of neurodegenerative diseases such as AD. Synapse dysfunction and loss are recognised as some of the earliest hallmarks of AD^{50–52}, occurring decades before the onset of clinical symptoms and closely correlates with the cognitive impairment associated with the disease^{53–56}. As a result, disruptions in synaptic communication can lead to connectivity and network dysfunctions, often manifesting as neuronal hyperexcitability⁵⁷. This hyperactivity not only disrupts normal neural activity but also contributes to a vicious cycle, further exacerbating synaptic damage through excitotoxicity and impaired plasticity⁵⁸. Over time, these synaptic disruptions and network imbalances worsen, eventually leading to widespread neuronal loss and cognitive decline.

1.1.3.3 The clinical phase

The clinical phase of AD begins when the cellular response can no longer maintain homeostasis, giving rise to symptomatic manifestations such as memory loss, disorientation, or impaired spatial perception. Early symptoms often include difficulty recalling recent events or lapses in judgment. As the disease advances, cognitive deficits expand to include impairments in language, visuospatial skills, and executive functions, alongside behavioural changes such as apathy, agitation, or delusions⁵⁹. In the most severe stages, individuals lose the ability to perform basic activities of daily living and require full-time care. This phase is accompanied by changes detectable through imaging, such as hippocampal atrophy, functional decline on MRI, and cerebrospinal fluid (CSF) alterations. Notably, biologically-defined AD, based on biomarker evidence, does not always align with clinically-defined probable AD, emphasising the complexity and heterogeneity of disease progression¹.

Clinical scales like the Mini-Mental State Examination (MMSE)⁶⁰ and Clinical Dementia Rating (CDR)⁶¹ are important tools for assessing disease severity and progression. The CDR provides a framework to categorise disease stages, ranging from preclinical AD with unimpaired cognition (CDR = 0) to mild cognitive impairment (MCI, CDR = 0.5), and further into stages of AD dementia: mild (CDR = 1), moderate (CDR = 2), and severe (CDR = 3). MCI, often considered a prodromal phase and concurrent to the onset of the clinical phase (**Figure 1-2**), is marked by measurable cognitive decline, particularly in memory, without significant impact on daily living. The transition from MCI to AD dementia (CDR > 1) is a critical inflexion point in disease progression. However, not all individuals with MCI progress to dementia; some remain stable, some progress at varying rates — either rapidly or slowly⁶². In other words, prodromal progression to AD dementia is remarkably heterogeneous^{63,64}. This is influenced by genetic factors (e.g., *APOE-ε4*, *TREM2*), comorbidities, cognitive reserve, and molecular heterogeneity^{64,65}.

These clinical scales are often complemented by biomarker-driven frameworks like the ATN (Amyloid, Tau, Neurodegeneration) classification⁶⁶. Integrating clinical assessments with biomarkers allows for a nuanced understanding of disease trajectories and for distinguishing AD from other dementias.

1.2 High-dimensional omics data

In the post-genomic era, agnostic computational biology analyses are increasingly used to generate disease mechanism hypotheses. This trend is accelerated by the growing availability of high-dimensional biological data powered by the emergence of high-throughput sequencing technologies. We can now capture high-dimensional data at each layer of biological systems, ranging from genomics, transcriptomics, and proteomics to metabolomics⁶⁷. In AD, omics experiments have been used in living patients and *post mortem* samples to capture molecular and cellular data across brain regions and pathology stages (**Figure 1-3**). This section provides an overview of omics layers, and the next-generation sequencing and mass spectrometry techniques used to acquire the multi-omics data analysed throughout this thesis, including genomics, transcriptomics, proteomics and epigenetics.

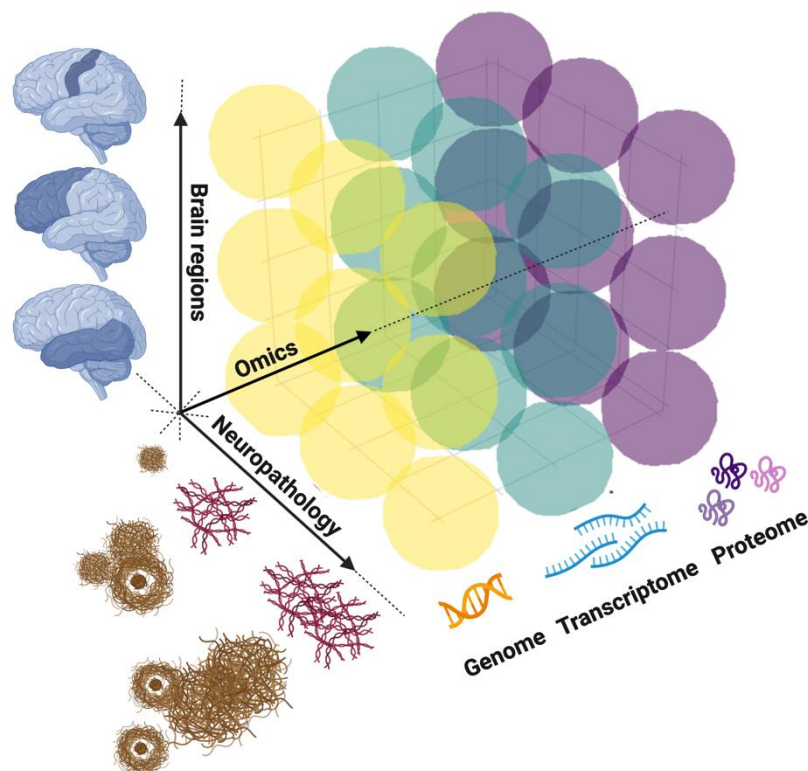


Figure 1-3 High-dimensional multi-omics profiling of post mortem samples in AD
Figure generated via [Biorender](#).

1.2.1 Genomics

Genomics quantifies the genetic architecture underlying biological systems. By analysing variations across thousands of genomes, researchers can identify genetic variations, such as Single Nucleotide Polymorphisms (SNPs), insertions and deletions, or copy number variants. The two broad categories of acquisition techniques covered in this thesis include: (1) genotyping, which is well-suited for detecting known SNPs and genetically stratifying individuals, and (2) whole-genome sequencing, which provides a more comprehensive view of genetic variation.

Genome-wide association studies (GWAS) have been instrumental in identifying genetic variants, both rare and common, associated with specific traits or diseases. These studies compare the frequency of variants in populations with a given trait to those without it, revealing variants significantly overrepresented in the affected population. GWAS findings also provide a foundation for calculating Polygenic Risk Scores (PRS).

PRS aggregate the effects of multiple SNPs to quantify an individual's genetic predisposition to a specific trait or disease. By summing the number of risk alleles an individual possesses, weighted by their effect sizes as determined by GWAS, PRS offer a cumulative measure of genetic risk. While each variant may contribute a small effect individually, the aggregation of multiple variants can yield substantial predictive power. PRS has proven especially useful for predicting the risk of polygenic traits like AD⁶⁸. Integrating PRS with other omics data, such as transcriptomics and proteomics, facilitates a deeper understanding of how genetic variation translates into molecular and cellular phenotypes⁶⁹.

1.2.2 Transcriptomics

Transcriptomics is the study of the complete set of Ribonucleic Acid (RNA) transcripts, including both coding and noncoding RNAs (e.g., mRNA, rRNA, tRNA, and ncRNA). Advances in high-throughput next-generation sequencing (NGS) have revolutionised transcriptomics by enabling the comprehensive analysis of RNA through RNA sequencing (RNA-seq). After tissue preparation, cDNA libraries are prepared by reverse transcription of

RNA into complementary DNA (cDNA), followed by adaptor ligation and amplification. These libraries are then sequenced using high-throughput platforms, such as Illumina sequencers, which amplify DNA on a flow cell and use fluorescently labelled nucleotides to read the sequence. After sequencing, the raw data is pre-processed through steps like quality control and alignment to a reference genome. RNA-seq can be performed at the bulk or single-cell level, as overviewed in **Figure 1-4**. Bulk RNA-seq quantifies the average gene expression in a given tissue, which provides a global picture of gene expression but masks cell-type specific signals. Single-cell RNA-seq (scRNA-seq) was developed to overcome these limitations, allowing the quantification of gene expression at the resolution of individual cells. scRNA-seq requires fresh tissues, limiting its application to *post mortem* brain samples. The development of single-nuclei RNA-seq (snRNA-seq) extended applications to frozen tissues like frozen brain samples by isolating nuclei instead of whole cells, as nuclei remain intact during freeze-thaw processes, whereas the cytoplasmic membranes of large cells like neurons are often damaged. The first snRNA-seq study of human *post mortem* AD samples was conducted by Mathys et al. in 2019, profiling the transcripts in more than 80,000 cells from 48 donors⁷⁰. This was a milestone in characterising cellular heterogeneity and uncovering cell type-specific differentially expressed genes in AD. Since then, transcriptomics studies have evolved to larger-scale analyses, genetically stratified cohorts, profiling of additional brain regions, and integration with multi-omics approaches.

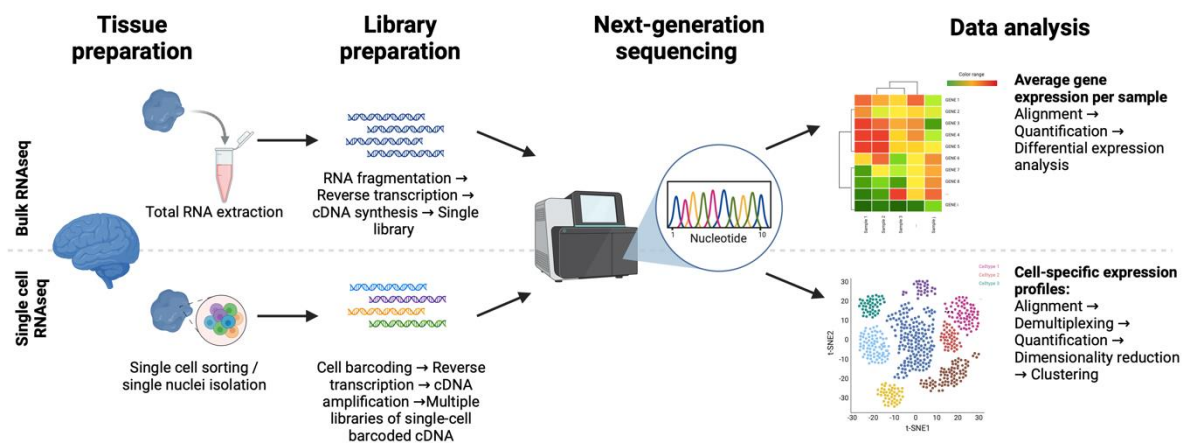


Figure 1-4 RNA-seq workflows
Figure generated via [Biorender](#).

1.2.3 Proteomics

Proteomics is the large-scale study of proteins, molecules translated from mRNA, which are the primary functional components of cells and are responsible for executing and regulating biological processes. Proteomics directly quantifies protein abundance in a sample (biofluids including CSF and plasma, or brain tissue), thereby providing a functional snapshot of cellular activity and insights into post-transcriptional regulation, protein modifications, and interactions. Mass spectrometry (MS) advances, particularly liquid chromatography-tandem mass spectrometry (LC-MS/MS), have enabled the high-throughput quantification of proteins in large-scale experiments. **Figure 1-5** provides an overview of the LC-MS/MS workflow, including pre-LC-MS/MS steps of tissue preparation, protein extraction and enzymatic digestion. Liquid chromatography (LC) separates peptides based on their hydrophobicity, allowing them to be analysed individually. The tandem mass spectrometry (MS/MS) step detects peptides based on their mass-to-charge ratio (m/z) and generates fragmentation patterns (MS/MS spectra) to identify peptides and infer proteins. By comparing these spectra to a reference database, modern proteomics workflows can now detect and quantify over 10,000 proteins in a single experiment⁷¹.

Proteomics profiling methods have substantially improved in recent years and offer untargeted mass-spectrometry techniques such as label-free or isobaric tagging approaches, both leveraged in this thesis. Label-free proteomics quantifies protein abundance directly from MS signal intensities, offering a cost-effective and flexible experimental design but with lower sensitivity to low-abundance proteins and greater variability between runs. Recent advancements, such as zenSWATH MS, have optimised label-free approaches to achieve deeper proteome coverage with minimal missing values, making it highly scalable for studies with many samples⁷². Isobaric tagging approaches, such as Tandem Mass Tag (TMT) proteomics, chemically label peptides to enable multiplexed analysis of multiple samples in a single run, providing room for higher scale experiments, higher sensitivity and reproducibility, though at the cost of more complex preparation. A landmark, large-scale bulk TMT proteomics study in 129 *post mortem* brains from AD and control donors was conducted by Seyfried et al. in 2017, uncovering co-expressed networks of proteins correlated with AD⁷³. While bulk proteomics remains a standard for untargeted experiments, more specialised proteomics approaches can be leveraged in specific biological contexts. For instance, synaptoproteomics

isolates proteins from synaptoneurosomes to focus on the molecular composition of pre- and post-synaptic compartments (**Chapter 4**). Additionally, imaging proteomics, including Imaging Mass Cytometry (IMC), enables the quantification of up to 40 protein markers alongside their spatial visualisation within a single tissue section, making it a great tool for studying spatially resolved protein abundance (**Chapter 5**).

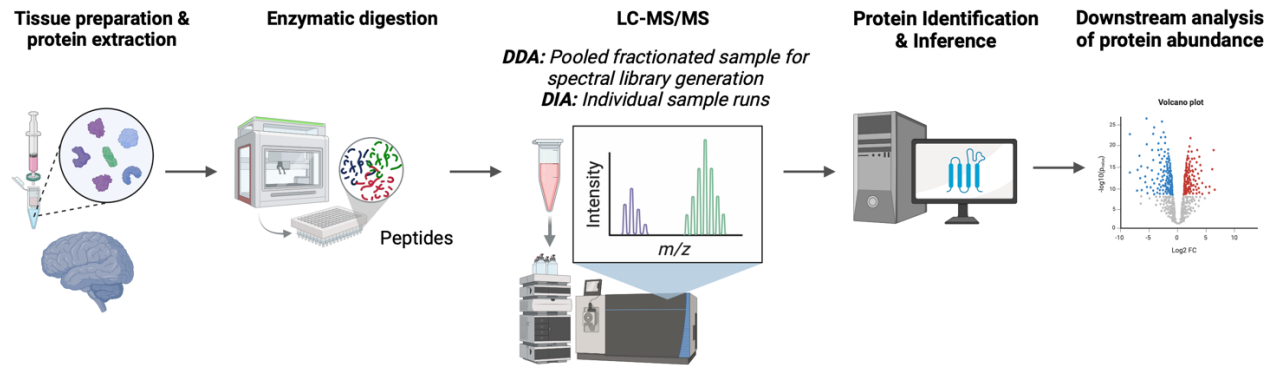


Figure 1-5 Proteomics workflow

Figure generated via [Biorender](#). DDA: Data-Dependent Acquisition. DIA: Data-Independent Acquisition.

1.2.4 Epigenetics

Epigenetics refers to the study of heritable changes in gene expression that occur without changes in the DNA sequence, mediated by molecular mechanisms such as DNA methylation or histone modifications (methylation/acetylation). These processes regulate the accessibility of the DNA to transcription factors (TF) and the transcriptional machinery, which modulate gene expression and cellular function. Chromatin Immunoprecipitation Sequencing (ChIP-seq) is a widely used technique to identify protein-DNA interactions. It enables the genome-wide mapping of TF binding sites and histone modifications by immunoprecipitating DNA bound to specific proteins or modifications, followed by sequencing. Tools such as ChEA3 facilitate TF enrichment analysis by integrating orthogonal omics data and leveraging a library of TF proteins and their target genes to predict key regulatory drivers of observed gene expression changes⁷⁴. Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) profiles chromatin accessibility by identifying regions of the genome that are open and accessible. This is measured by the ability of the transposase enzyme Tn5 to cut accessible regions of chromatin and insert adapters for sequencing. These accessible regions, visualised

as peaks in sequencing data, often correspond to active regulatory elements such as promoters and enhancers. Emerging technologies like the 10x Multiome platform now enable simultaneous profiling of chromatin accessibility and gene expression at the single-cell level⁷⁵.

1.3 Integrative analyses

1.3.1 What is multi-omics integration?

Omics datasets reflect quantitative levels (i.e. gene or protein expression levels) and composition (sequence variations, isoforms, or molecular subtypes) of biomolecules within a biological system. Single-platform approaches attempt to model the relationship between one omic layer and the phenotype, which often consists in identifying differential features associated with the disease. These can be useful disease markers and provide functional insights into dysregulated biological pathways at play. Nevertheless, analysing only one omic layer often provides a partial view of biological systems. Single omic analyses overlook the interplay between biological layers and thus can provide a simplistic interpretation of the dysregulations in disease. Instead, integrating different omics data types helps to model the interplay between molecular layers and thus may provide improved analytical value for gaining full mechanistic insights into diseases. Hence, the translational value of omic data lies in their integration and joint analysis, a field referred to as multi-omics. The Cancer Genome Atlas⁷⁶ paved the way for many seminal studies in the multi-omics field. The field is now expanding towards all disease areas, including neurodegenerative, and addressing a range of research questions from disease subtyping^{77–79}, molecular mechanisms discovery⁸⁰, and biomarkers for diagnostics or prognostics⁸¹.

1.3.2 Multi-omic integrative analyses in human AD

Moving beyond single-omic approaches to multi-omic integrative analyses is crucial if we are to build up a full molecular profile of the cellular changes that occur in AD. The availability of large-scale multi-omics cohorts, such as ROSMAP, MSBB, Mayo or the Multi-Omics Atlas Project (MAP)^{82–84}, are facilitating the generation of new biological hypotheses of AD based on different omic profiles, brain regions and disease stages. The field is transitioning away

from genetics-only focused studies, including GWAS, or transcriptomics focused, towards integrating multiple omics datasets to uncover the molecular drivers of AD. This shift highlights the fact that no single omic modality on its own can capture all the disease-related changes underlying AD. To this date, several patient-specific multi-omics integrative analyses to study human AD have been published, and these were summarised in **Table 1-1**. Below, I explore the commonalities, differences and gaps in these studies. For each recorded studies, the following were summarised: study design and sample size, integrated omics, AD definition, main findings, integration objectives and integration method. Exclusion criteria are specified where applicable. Data presented in the recorded studies were specifically generated unless explicitly referenced as originating from a publicly available AD cohort. Consequently, public datasets—including those from ROSMAP, MSBB, ADNI, and others—may exhibit partial overlap in participant samples.

From the recorded studies, transcriptomics emerges as the most widely used omic type (13/15, ~86%), followed by proteomics (9/15, ~60%). This aligns with broader findings in the literature, where the integration of transcriptomics and proteomics is the most common combination in non-cancer-focused studies⁸⁵. As mRNA is translated into protein, it is usually assumed that mRNA levels and protein should have a high correlation. However, most studies that investigated pairwise mRNA and protein correlation find only a weak correlation (~ 0.3)^{86,87}, indicative of complex post-translational regulatory mechanisms. The integration of transcriptome and proteome profiles offers an opportunity to reveal distinct and complementary features, which ultimately increase the coverage and statistical power of pathway analyses.

While the recorded studies differed widely in their specific objectives, these all grouped into three broad integration aims: (1) biomarker discovery, (2) understanding molecular mechanisms, and (3) identifying molecular subtypes of AD. Biomarker discovery was the predominant focus among the recorded studies, including multi-omic biomarkers derived from blood, plasma, CSF, or brain tissues to differentiate AD patients from controls and predict disease stages. All studies were cross-sectional, which is expected for *post mortem* work but surprising for *ante mortem* studies, given the potential for longitudinal sampling with blood and CSF. This limitation, often linked to the high costs of multi-omics studies, restricts insights into the temporal dynamics of AD progression. Furthermore, even in *post mortem* studies, omics data integration with lifetime longitudinal cognitive or clinical data remains

underexplored. Also, few biomarker studies also focused on mechanistic insights to improve their translational potential for therapeutic applications. In fact, only one study covered all three integration aims into a unified framework⁶⁵. Most studies relied solely on clinical or pathological definitions of AD rather than stratifying patients based on a combination of both. Finally, since most studies used bulk omics data, findings were not resolved to cell-specific mechanisms. Given that the cellular phase represents a crucial therapeutic window in AD, a lack of focus on glial mechanisms remains a critical gap.

The computational methods for multi-omics integration varied widely, indicating it's not a one-size-fits-all endeavour. Statistical regression methods, such as Sparse Partial Least Squares (sPLS) and LASSO, were used for feature selection and biomarker discovery. Unsupervised learning methods like MOFA, iClusterBayes, and Non-Negative Matrix Factorization (NMF) can uncover hidden patterns in the data, and these were employed to identify molecular subtypes. Supervised learning methods, including machine learning classifiers, are widely applied for disease prediction. Finally, network-based integrative approaches like Similarity Network Fusion (SNF) and CellChat provide insights into molecular and intercellular interactions. This highlights the need for integrated multi-omics workflows that combine multiple modelling approaches and allow researchers to fully capture the complexity of AD.

<i>Study (year)</i>	<i>Design, Sample size</i>	<i>Omics</i>	<i>AD definition</i>	<i>Main findings</i>	<i>Integration method</i>	<i>Integration aim(s)</i>	<i>Ref.</i>
<i>Multi-omics signature of brain amyloid deposition in asymptomatic individuals at-risk for Alzheimer's disease: The INSIGHT-preAD study (2019)</i>	Cross-sectional <i>ante mortem</i> Blood, plasma n=48 amyloid positive n=48 amyloid negative	Blood RNA-seq, plasma metabolomics, lipidomics	Neuropathological diagnosis of AD (Amyloid positivity)	The model's top metabolites and transcripts represented the most discriminant omics features. 99.4% prediction accuracy for amyloid positivity.	Sparse Partial Least Squares Discriminant Analysis	Biomarkers	88
	Neither sex nor age significantly differed between the amyloid groups				Correction for sex and <i>APOE</i> genotype		
<i>Identification of functionally connected multi-omic biomarkers for Alzheimer's disease using modularity-constrained Lasso (2020)</i>	<i>Post mortem</i> Brain n=115 cognitively normal, n=67 MCI, n=80 AD	Genotype, transcriptomics, proteomics	Clinical diagnosis of AD	276 multi-omic biomarkers, including SNPs, genes and proteins, were identified to bear predictive power. Genetic variants and cascade effects on the transcriptome and proteome levels could explain cognitive performance deterioration in AD patients.	Modularity constrained LASSO	Biomarkers	89
	Subjects with missing data in one or more data types were excluded				Adjusted to remove the effect of age, sex and education		
<i>Bayesian integrative analysis of epigenomic and transcriptomic data identifies Alzheimer's disease candidate genes and networks (2020)</i>	<i>Post mortem</i> Brain AD n=141 Controls n=92	Transcriptomics, Epigenomics (Histone acetylation, DNA methylation)	Combination of clinical and neuropathological diagnosis of AD	Detected 816 genes with consistent differences between persons with AD and controls.	Gene-centric integrative coefficient: Bayesian hierarchical model is used to study the distribution of the integrative coefficient	Biomarkers	90
	Age and sex matched						
	Donors with mild cognitive impairment were excluded						
<i>Epigenomics and lipidomics Integration in Alzheimer's Disease: Pathways Involved in Early Stages (2021)</i>	<i>Post mortem</i> Plasma MCI n=22, Controls n=5	Epigenomics, lipidomics	Combination of clinical and neuropathological diagnosis of AD	25 miRNAs and 25 lipids with higher loadings in the sPLS regression were selected and associated with early AD. Fatty acid mechanisms are proposed to be involved in early disease pathogenesis.	Sparse Partial Least Squares	Biomarkers, Molecular mechanism	91
	Patients with known major neurological or psychiatric conditions were excluded						
<i>Multionics integrative analysis identifies APOE allele-specific blood biomarkers associated to</i>	<i>Post mortem</i> Plasma and brain	GWAS, differential mRNA and protein expression analyses,	Clinical diagnosis of AD + excluded controls positive	Identification of APOE-specific signatures; identified genes and pathways contributing to AD in	Robust Rank Aggregation	Biomarkers	92

<i>Alzheimer's disease etiopathogenesis (2021)</i>	Meta-analysis combining ADDN, ADNI, MAYO, ROSMAP, MSBB (n=10,551)	single-nuclei transcriptomics	for amyloid pathology	both APOE dependent and independent fashion.			
	Individuals of non-European ancestry were excluded						
	Sex stratified meta-analysis						
<i>Joint Deep Semi-Non-Negative Matrix Factorization for Learning Integrative Representation of Molecular Signals in Alzheimer's Disease (2021)</i>	<i>Ante mortem</i> Blood AD n=95 Controls n=199 Cohorts include: ANM consortium, ADNI	Transcriptomics, epigenomics (DNA methylation),	Clinical diagnosis of AD	Identified AD and age-related modules from the latent matrices.	Joint Deep Semi-Non-Negative Matrix Factorization	Biomarkers	93
					Adjusted to remove the effect of sex differences		
<i>An integrative multi-omics approach reveals new central nervous system pathway alterations in Alzheimer's disease (2021)</i>	Cross-sectional <i>ante mortem</i> CSF n=120 with normal, MCI or mild AD	Proteomics, metabolomics, lipidomics	Combination of clinical and neuropathological diagnosis of AD	Novel molecular pathway alterations associated with AD pathology, including hemostasis, immune response, and extracellular matrix signalling	Multi-Omics Factor Analysis	Molecular mechanism, Biomarkers	94
					Adjusted to remove the effect of age, sex and education		
<i>Identification of Blood-Based Multi-Omics Biomarkers for Alzheimer's Disease Using Firth's Logistic Regression (2022)</i>	Cross-sectional <i>ante mortem</i> Blood n=32 AD n=15 non-AD subjects	16 transcript genes, 14 metabolites, and 9 proteins (cytokines).	Clinical diagnosis of AD	The model revealed 18 potential biomarkers of AD, consisting of seven metabolites, two transcriptomes, and nine cytokines.	Logistic regression model	Biomarkers	95
<i>Alzheimer's disease prediction based on continuous feature representation using multi-omics data integration (2022)</i>	<i>Post mortem</i> Brain n=257 cognitively normal n=440 AD samples	Transcriptomics, epigenomics (DNA methylation)	Clinical diagnosis of AD	AD can be identified with 90% prediction accuracy.	ML classifiers	Biomarkers	81
	Cohort(s) include: ROSMAP						

<i>Multi-Omics-Based Autophagy-Related Untypical Subtypes in Patients with Cerebral Amyloid Pathology (2022)</i>	Cross-sectional <i>ante mortem</i> Blood-derived biopsied specimens n=170 patients with cerebral amyloid pathology Cognitively normal n=40 MCI n=65 AD n=65	Genotyping, miRNA transcriptomics, proteomics (TMT-based blood proteomics), immunoassays	Clinical diagnosis of AD for patients with cerebral amyloid pathology	Identification of subtypes and potential association of subtypes with autophagy pathways.	Multi-Omics Factor Analysis	Subtyping	96
<i>Multi-omic integration via similarity network fusion to detect molecular subtypes of ageing (2023)</i>	<i>Post mortem</i> Brain n=1313 (838 AD, 475 controls)	Transcriptomics, epigenomics (DNA methylation, histone acetylation), proteomics, metabolomics	Neuropathological diagnosis of AD	Identified subtypes associated with cognitive decline.	Similarity network fusion Adjusted to remove the effect of age, sex and education	Subtyping	97
<i>Integrated multi-omics analysis of Alzheimer's disease shows molecular signatures associated with disease progression and potential therapeutic targets (2023)</i>	Cross-sectional <i>post mortem</i> and <i>ante mortem</i> samples of brain, blood, and CSF. 5 GEO transcriptomic datasets (n=1,311), Blood proteomics (n=60), Brain proteomics (n=90), CSF proteomics (n=40) Cohort(s) include: HBTRC, ROSMAP, ADRC, MSBB, Shandong Provincial Hospital	Transcriptomics, proteomics, metabolomics	Defined in respective studies of meta-analysis	Deregulated pathways included neurotransmitter signalling, oxidative stress, complement cascade, and metabolic remodelling. The commonality in pathways between human and mouse datasets highlighted conserved disease mechanisms.	Pathway enrichment analysis (Enrichr, MetaboAnalyst), commonality analysis	Biomarkers	98
<i>Unravelling the intercellular communication disruption and key pathways in Alzheimer's disease: an integrative study of single-nucleus transcriptomes and genetic association (2024)</i>	<i>Post mortem</i> Brain n=48 (24 AD, 24 controls) Sex/age of death-matched	GWAS, Genomics (WGS), Transcriptomics (snRNA-seq),	Neuropathological diagnosis of AD	Identified 190 dysregulated ligand-receptor interactions and 44 significant GO terms (such as amyloid precursor protein processing and calcium ion transport), highlighting disrupted cell-cell communication. Drug	CellChat for intercellular communication; MAGMA and PRSet for pathway analysis; NicheNet	Molecular mechanisms	99

	Cohort(s) include: ROSMAP			repurposing identified 9 FDA- approved drugs targeting dysregulated LR genes.	for ligand-target gene predictions	
<i>A comprehensive multi-omics analysis reveals unique signatures to predict Alzheimer's disease (2024)</i>	<i>Post mortem</i> Brain n=455 participants (307 AD, 148 controls)	Genomics (SNP), Transcriptomics, Epigenomics (CpG methylation), proteomics	Clinical diagnosis of AD	Integration improved predictive accuracy, identifying robust biomarker panels compared to individual omics platforms alone.	Sparse generalized canonical correlation analysis	Biomarkers
						100
<i>Brain high-throughput multi-omics data reveal molecular heterogeneity in Alzheimer's disease (2024)</i>	<i>Post mortem</i> Brain n=278 (255 AD, 23 controls)	Transcriptomics, proteomics, metabolomics, lipidomics	Clinical diagnosis of AD	Identified 4 molecular profiles, and one associated with worse cognitive outcomes, faster disease progression, and dysregulated pathways like synaptic signalling, endocytosis, and mTOR signalling. CSF biomarkers were identified for disease staging and monitoring progression.	iClusterBayes Adjusted for sex, age at death	Biomarkers, Subtyping, Molecular mechanisms
						65
	Cohorts include: Knight ADRC, MSBB, ROSMAP					

Table 1-1 Literature review of integrative multi-omic analyses in human AD

Integrative analyses in human AD were sourced from PubMed by applying the following search criteria: “Multi-omics” AND/OR “Integrative” AND “Alzheimer’s”. I excluded studies that were (1) not patient-specific, (2) didn’t explicitly perform a joint integration method, or (3) those based on animal models.

1.4 Thesis overview

Using multi-omics integrative analyses, I explore how glial activation, specifically, of microglia and astrocytes, are active players in AD pathogenesis. Through a bottom-up approach, this research first develops an integrative transcriptomics-proteomics integration pipeline to generate hypotheses on AD mechanisms (Chapter 2). This is followed by investigations into the relationships between glial activation and cognitive decline (Chapter 3); and between reactive astrocytes and synaptopathy (Chapter 4). Finally, these findings are validated at the spatial level using imaging mass cytometry (Chapter 5). The overall hypothesis is that glial activation contributes to cognitive decline through its effect on synaptopathy, forming a glia-cognition link that underlies the observed clinical heterogeneity in AD (**Figure 1-6**). Research aims, chapter-specific hypotheses and knowledge gaps addressed are outlined below.

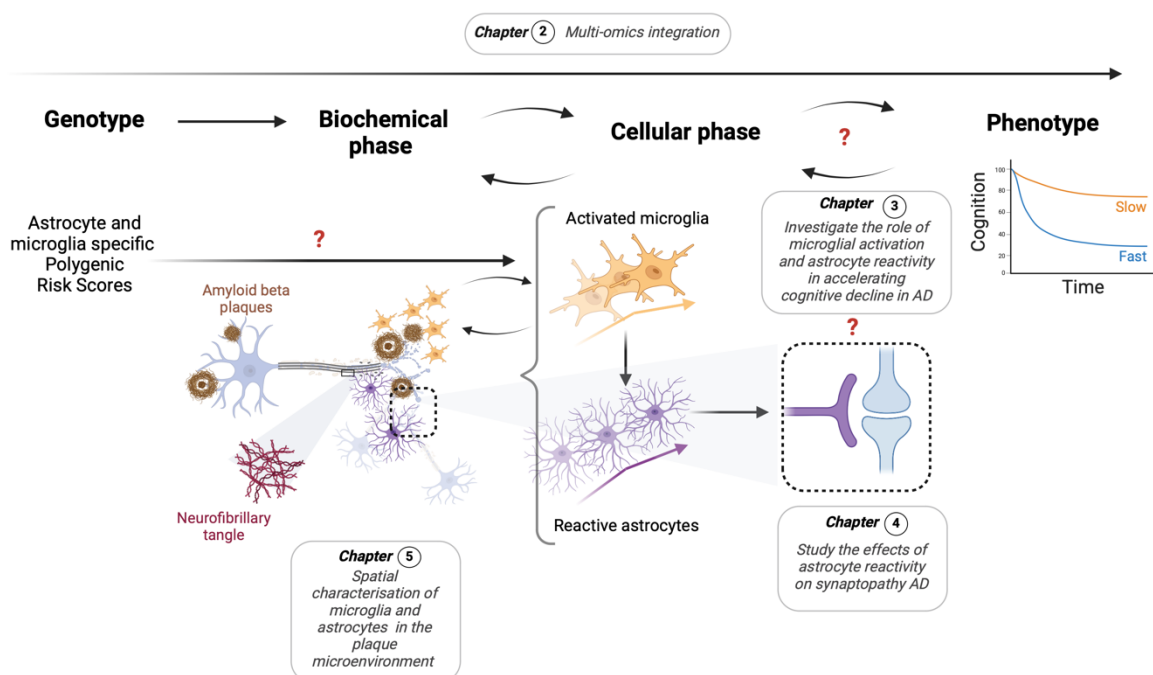


Figure 1-6 Overview of thesis: Multi-omics integration in AD: Linking genotype to phenotype through glial activation, synaptopathy and cognitive decline

The effect of the environment (i.e. modifiable risk factors) is not visualised here. Figure generated via [Biorender](#).

Aim 1: Develop and validate a transcriptomics-proteomics integration pipeline

In *Chapter 2 “Omix: A Transcriptomics-Proteomics integration pipeline”*, I will develop and validate *Omix*, a bioinformatics pipeline for end-to-end transcriptomics-proteomics integration, including an R package and reproducible workflows. The pipeline will be tested on public gold-standard AD datasets to uncover molecular mechanisms underlying AD pathology and assess the added value of multi-omics over single-platform analyses.

- **Hypothesis:** Integrating transcriptomics and proteomics will provide more accurate, predictive and robust disease-specific molecular signatures compared to single-omics analyses.
- **Knowledge gaps addressed:**
 - a) **Integration challenges:** Workflow capable of combining diverse modelling techniques to address multiple integration objectives and capture the complexity of AD.

Aim 2: Characterise glial and synaptopathy mechanisms associated with AD progression and clinical heterogeneity

This aim is divided into two parts:

Chapter 3 “Integrative multi-omics analysis of glial signatures associated with accelerated cognitive decline in AD”: Using multi-omics analyses, I will investigate the role of microglial activation and astrocyte reactivity in accelerating cognitive decline in AD. I will mechanistically relate this to specific subpopulations of microglia and astrocytes, identify their molecular signatures, and assess their genetic enrichment. Finally, I will validate brain molecular signatures by assessing their relevance in CSF and using them for prediction tasks.

Chapter 4 “A synaptic-astrocytic proteomic signature associated with synaptopathy in AD”: I will focus on the tripartite synapse and study the effects of astrocyte reactivity on synaptic dysfunction and loss in AD.

- **Hypothesis:** Glial activation contributes directly and indirectly to clinical heterogeneity in AD through their implication in neuroinflammation and synaptopathy. These processes are central to the differential progression of slow- versus fast-progressing AD phenotypes.
- **Knowledge gaps addressed:**
 - a) **Cell subpopulation-specific dynamics:** Explore the molecular and functional diversity of activated microglia and astrocyte subpopulations and their role in synaptopathy.
 - b) **Heterogeneous progression:** Investigate how glial activation and synaptopathy contribute to the differential trajectories and clinical heterogeneity of AD.
 - c) **Biomarkers for stratification:** Related signatures in CSF and plasma may provide biomarkers to identify patients at higher risk for rapid cognitive decline.
 - d) **Resolve astrocyte changes at the synapse level:** Use synaptic proteomics to identify these changes at the synapse level (i.e. the tripartite synapse).

Aim 3: Spatial validation of microglia and astrocyte subpopulations according to the plaque microenvironment

In *Chapter 5 “Spatial omic characterisation of glial activation in AD using imaging mass cytometry”*, findings from *Aim 2* will be validated at the spatial level using Imaging Mass Cytometry (IMC). Through this approach, I will assess the spatial dynamics of microglia and astrocyte subpopulations relative to the amyloid plaque microenvironment in AD.

- 1) **Hypothesis:** Activated glial subpopulations exhibit distinct spatial patterns and functional roles depending on their proximity to amyloid plaques, influencing the local neuropathological environment and disease progression.
- 2) **Knowledge gaps addressed:**
 - a) **Spatial profiling:** Mapping the spatial distribution of glial subpopulations and their interactions with the surrounding neuropathological environment.

1.5 Permissions information chapter 1

<i>Page No.</i>	<i>Type of work:</i>	<i>Name of work</i>	<i>Source of work</i>	<i>Copyright holder</i>	<i>permission requested on</i>	<i>I have permission yes /no</i>	<i>Permission note</i>
21	figure	Figure 3. The genetic landscape of Alzheimer's disease	Alzheimer's disease, Scheltens, Philip et al. (2021) The Lancet, Volume 397, Issue 10284, 1577 - 1590	Elsevier	23rd Nov 2024	yes	Permission granted to reproduce the figure. License Number 5914771428335
22	figure	Figure 1. Timing of Major AD Pathophysiological Events in Relation to Clinical Course	Alzheimer Disease: An Update on Pathobiology and Treatment Strategies, Long & Holtzman (2019) from Cell, Volume 179, Issue 3, 312 – 339	Elsevier	23rd Nov 2024	yes	Permission granted to reproduce the figure. License Number 5914790029341

Chapter 2 *Omix*: A Transcriptomics-Proteomics integration pipeline

Chapter declarations:

1. Some of the work in this chapter has been previously made available as a preprint:

El onore Schneegans, Nurun Fancy, Michael Thomas, Nanet Willumsen, Paul Matthews and Johanna Jackson (2023). “Omix: A Transcriptomics-Proteomics Integration Pipeline”. In: *BioRxiv*. doi: [10.1101/2023.08.30.555486](https://doi.org/10.1101/2023.08.30.555486).

The copyright holder for this preprint is **Eléonore Schneegans**.
[CC-BY-NC-ND 4.0 International license](#).

2. The Omix pipeline was awarded the 2023 UK DRI Computational Reproducibility Prize.

3. Some of the work in this chapter has been previously written in my Early Stage Assessment (October 2022).

4. I thank the following contributors:

- Nurun Fancy: *Omix* pipeline contributor.
- Michael Thomas: Tested the *Omix* pipeline (Installation and running the vignettes); Processed and annotated 10x Multiome data used in validation.

All the internal data used for analysis was generated entirely by the wet lab team, specifically:

- E.A led the IHC and snRNA-seq data generation.
- N.W led the IHC and snRNA-seq data generation.
- M.P snRNA-seq & IHC data generation.
- V.C contributed to IHC data generation.
- D.C contributed to IHC data generation
- R.C.J M contributed to IHC data generation.
- H.W contributed to proteomics data generation.
- R.Y contributed to proteomics data generation.
- A.MG contributed to bulk transcriptomics data generation.
- X.Z contributed to bulk transcriptomics data generation.

Abstract

Background: Multi-omics integration uncovers the relationships between molecular layers, providing comprehensive mechanistic insights into health and disease. A critical aspect of this landscape is the integration of transcriptomics and proteomics, which can facilitate mapping gene expression networks and protein activity, revealing key regulatory nodes that become dysregulated in disease conditions. A specialised integration workflow for bulk transcriptomics and proteomics, the most frequently recorded omics combination in non-cancer-focused studies, has not been documented.

Objectives: My research addresses topics linking multi-omics integration and dementia research, aiming to gain mechanistic insights into AD neuropathogenesis and progression. Using a bottom-up approach, I first build an integrative pipeline to generate new hypotheses on AD mechanisms.

Methods: I have built *Omix*, a specialised R package for pre-processing, analysing, integrating, and visualising bulk transcriptomics and proteomics data. By embedding state-of-the-art bioinformatics tools with novel features into a user-friendly wrapper, *Omix* streamlines the integrative analysis of bulk transcriptomics and proteomics data, thereby facilitating complex biological investigations, from biomarker discovery to patient stratification. *Omix* was used on a public AD dataset to analyse patient-specific transcriptomic and proteomic data. Initial validation has included single omic analysis using standard single platform models, followed by integration, and comparison of the molecular signatures derived in both approaches.

Results: Integrative analyses were used to identify molecular signatures of AD pathology and were functionally annotated to define corresponding molecular mechanisms. Differential analyses revealed partial concordance between transcriptomics and proteomics. Comparative analyses highlight the value of taking an integrated multi-omics approach over single omics as they partially overlap with single platform signatures and provide complementary insights. In a case study, I demonstrate *Omix*'s capability to generate a pseudotemporal disease profile, providing valuable insights into potential regulatory pathways by identifying key transcription factor proteins, mainly STAT1, STAT3 and CEBPB, that regulate microglia gene networks in the disease state. I validated these findings using single nuclei multiomics data (snATAC-seq+snRNAseq) to confirm that the identified transcription factor proteins exhibit differentially accessible chromatin regions at their target sites. These regions were linked to genes within the identified microglia network, further supporting their regulatory role in AD.

Conclusions: *Omix* is an R package that can be applied agnostically to integrate bulk transcriptomics and untargeted proteomics datasets. It empowers both experts and non-experts to harness the full potential of end-to-end joint transcriptomics-proteomics analysis in a reproducible fashion, thereby enabling advances in integrative multi-omics research.

Graphical abstract

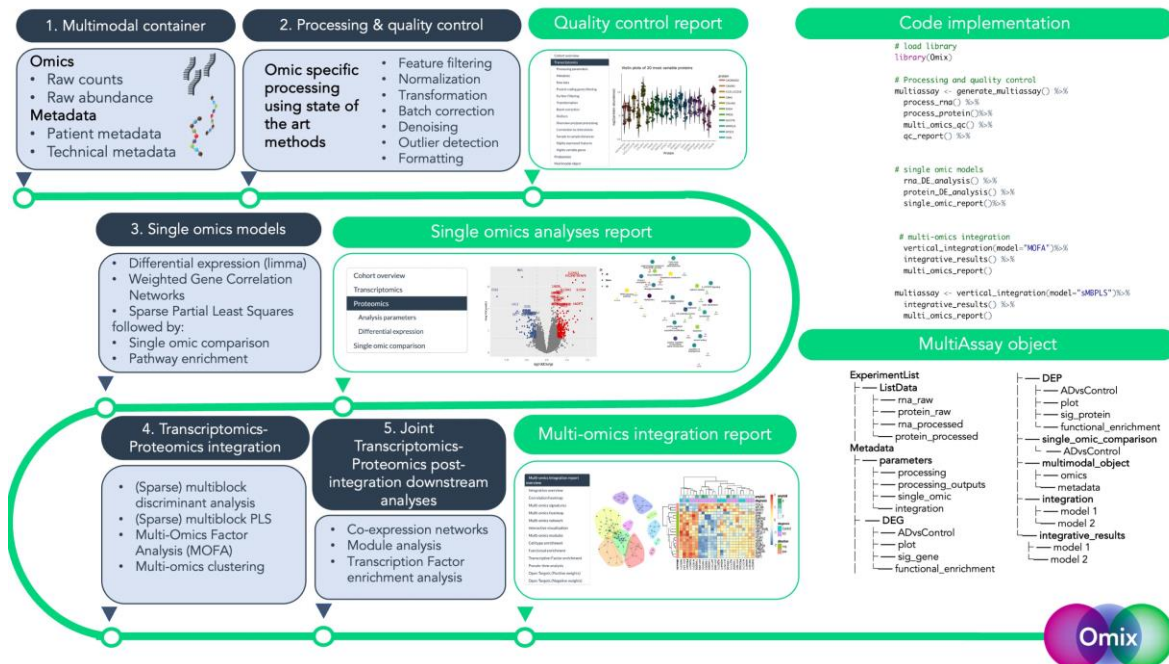


Figure 2-1 The Omix pipeline

Omix is a wrapper tool organised around five layers: (1) preparation of the multimodal container, (2) data processing and quality control, (3) single omic analyses, and (4) Transcriptomics-Proteomics vertical integration, (5) Joint Transcriptomics-Proteomics post-integration downstream analyses, producing standardised outputs that can be visualised in interactive HTML reports.

2.1 Introduction

Recent advancements in high-throughput omics techniques have transformed biomedical research by enabling the comprehensive capture of biological information across multiple molecular layers. This progress has led to pioneering discoveries in understanding biological systems. However, analysing individual omic layers offers only a partial insight into the complexity of these systems. Multi-omics integration uncovers the relationships between molecular layers, providing comprehensive mechanistic insights into health and disease⁶⁷.

2.1.1 From single platform to integrative multi-omics analyses

Omics datasets reflect the abundance or structural characterisation of biomolecules involved in biological processes. Single-platform approaches attempt to model the relationship between one omic layer and the phenotype, thereby identifying differential features associated with the disease. These can be useful disease markers and provide functional insights into which biological pathways may dysregulate in disease. Nevertheless, analysing only one omic layer often only provides a partial view of the biological system. Single omic analyses overlook the interplay between biological layers and thus can provide a simplistic interpretation of the dysregulations in disease⁶⁷. Integrating different omics data types helps to model the relationships between molecular layers and thus may provide improved analytical value for gaining full mechanistic insights into diseases. Hence, the translational value of omic data lies in their integration and joint analysis, a field referred to as integrative multi-omics⁶⁷. The Cancer Genome Atlas paved the way for the plethora of studies that emerged in the multi-omics field^{77,101}, and the field is now expanding towards all disease areas, addressing a range of research questions from disease subtyping, biomarker & molecular mechanisms discovery to classification tasks for diagnostics or prognostics.

The term integrative analysis has been used interchangeably to describe both horizontal and vertical integration^{102–104}. Horizontal integration focuses on a single type of omic data acquired across patient cohorts or groups. Using features (i.e. genes, proteins, SNP) as anchors can help correct for technical confounders and other batch effects. Instead, vertical integration used in multi-omics analyses is patient-specific and enables the use of individuals as anchors in the integrative process. In a joint integration setting, the goal is to find major sources of variation

and identify sample structures based on multiple molecular layers, thereby providing a joint picture of patient-to-patient (or cell-to-cell where in a single-cell framework) heterogeneity. From a modelling perspective, this requires sophisticated modelling approaches. No single method can address all research questions, and researchers should tailor the modelling approach based on their research goal and the omic layers considered for integration.

Overall, joint vertical integrative methods broadly fall within (1) clustering, (2) network, (3) correlation/factor analysis approaches, summarised in **Figure 2-2**. The correlation/factor analysis category comprises a vast number of both supervised and unsupervised models. Supervised models such as DIABLO¹⁰⁵ decompose data sets into components to help identify driving omic features responsible for the relationship between omics layers and the clinical phenotype. Unsupervised multi-omics integration methods like Multi-Omics Factor Analysis (MOFA)¹⁰⁶ also look for patterns within and across data types by inferring latent components, though this time in a label-agnostic fashion, i.e. without knowledge of the identity or label of the analysed samples. This category of models can be used to derive multi-omics signatures of disease and characterise molecular mechanisms, which is my research goal of interest. Meanwhile, clustering and network approaches have been used mostly for disease subtyping and classification tasks^{107–109}.

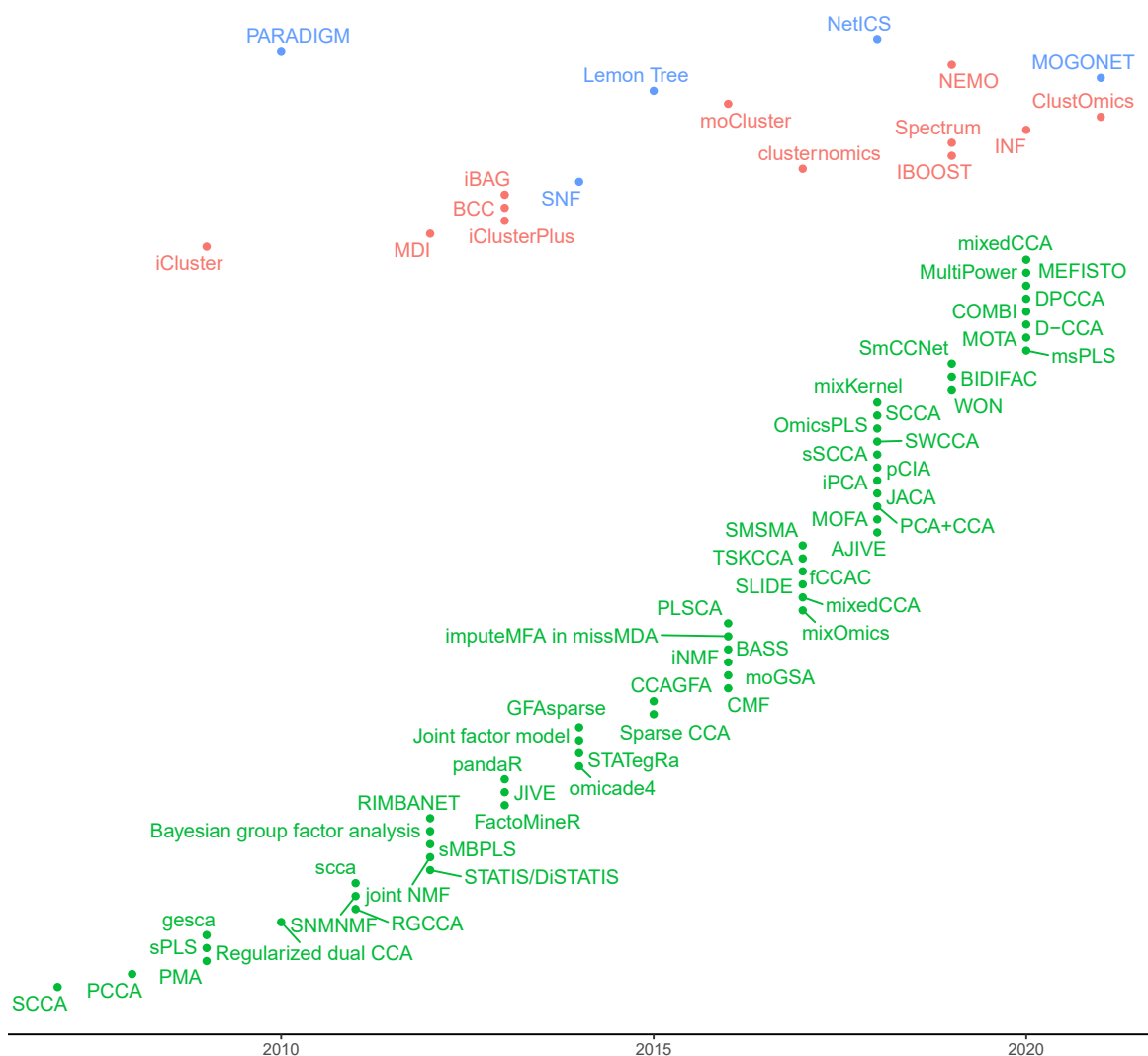


Figure 2-2 The landscape of vertical integrative models

Vertical integration models recorded from the literature. Different types of methods include multi-omics clustering methods (red), multi-omics correlation or factor analysis methods (green) and multi-omics networks (blue).

2.1.2 Transcriptomics-proteomics integration

A critical aspect of this landscape is the integration of transcriptomics and proteomics, the most frequently integrated omics combination in non-cancer-focused studies⁸⁵. Despite the central dogma of molecular biology suggesting a direct correlation between mRNA and protein levels, the reality is more complex. Transcriptomics, which reveals mRNA levels as precursors to protein synthesis, often does not correlate directly with protein levels¹¹⁰. This discrepancy is due to post-transcriptional modifications, regulatory mechanisms, and differing half-lives¹¹¹. Proteins, being closer to the phenotype and with a more stable structure than transcripts, offer

insights that are often more directly translatable for drug development. Proteomics can detect changes in low-abundance proteins crucial in disease processes, which transcriptomic assays might overlook due to their different sensitivity levels. Conversely, transcriptomic techniques can identify a broader range of transcripts than proteomics. Therefore, employing both modalities could offer a more comprehensive understanding of disease mechanisms due to their complementary¹¹². In addition, proteins like transcription factors (TF) directly influence gene expression, cellular states, and overall biological functions. Therefore, integrating transcriptomics with proteomics can facilitate mapping gene expression networks and protein activity to reveal key regulatory nodes that become dysregulated in disease conditions¹¹², and ultimately will help improving the identification of biomarkers and novel therapeutic targets.

2.1.3 Integration challenges

Integrating multiple omics data layers poses significant challenges, including for data processing, modelling with diverse signal/noise, the biological interpretation of results, as well as issues related to data annotation, storage, and computational resources¹¹³. The complexity inherent in omics data necessitates the use of diverse processing methods, often requiring the combination of multiple software tools and extensive computational expertise. More so, the rapid evolution of integration tools introduces obstacles to reproducibility and hampers the achievement of FAIR principles (Findability, Accessibility, Interoperability, Reusability) of multi-omics efforts¹¹⁴. The field of multi-omics research has seen the development of various integration pipelines, such as Miodin, MiBiOmics, Muon, and Movics^{115–118}. These tools, function as wrapper packages, supporting various omics types and incorporate a range of integrative models, from network-based to clustering approaches.

While multi-omics analyses bring new research opportunities, they also introduce new management, processing and analysis bottlenecks¹¹⁹. Given the diversity of integrative analysis methods, there is a growing need for end-to-end bioinformatics pipelines. Only with such frameworks in place will adoption of multi-omics analyses grow, and will researchers be able to unlock the full potential of multi-omics data.

2.1.4 Aim and objectives

In integrative multi-omics analyses, the primary goal is not merely to integrate multi-omics data, but also to gain insights that enhance our understanding of the biological processes and regulatory mechanisms at play. There is a pressing need for omic-specific workflows that process, integrate and guide users through curated pre- and post-integration steps. A specialised integration workflow for bulk transcriptomics and proteomics, the most frequently recorded omics combination in non-cancer-focused studies⁸⁵, has not been documented, highlighting a critical gap in multi-omics research.

In this Chapter, I aim to build *Omix*, a transcriptomic and proteomics integration and analysis bioinformatics pipeline implemented as an R package. The goal is to demonstrate that integrating multiple omics layers, here the transcriptome and proteome, can provide a more accurate, reproducible, and systematic understanding of the molecular and regulatory mechanisms underlying AD pathology compared to single-platform analyses.

To address this hypothesis, the following objectives have been set:

1. Build an R package suited to the processing, integration and downstream analyses of bulk transcriptomics and proteomics data.
2. Test the pipeline on publicly available AD datasets to compare molecular signatures derived from single-platform and multi-omic analyses.
3. Test the pipeline on an internal bulk transcriptomics-proteomics to characterise TF-regulated AD modules and validate whether the predicted regulatory relationships from the integrated bulk analysis are supported by changes in snATAC-seq + snRNA-seq data.

2.2 Omix pipeline implementation

2.2.1 Package Overview

The *Omix* pipeline offers an integration and analysis framework intended to flexibly pre-process, analyse, and visualise transcriptomics-proteomics data to address research questions¹²⁰. *Omix* is a wrapper tool built on five consecutive blocks: (1) preparation of the

multimodal container, (2) data processing and quality control, (3) single omic analyses, and (4) Transcriptomics-Proteomics vertical integration, (5) Joint Transcriptomics-Proteomics post-integration downstream analyses, presented in **Figure 2-1**.

2.2.2 Multimodal container

The core *Omix* pipeline is organised around wrapper functions that take and return a MultiassayExperiment object, a widely recognised Bioconductor format for managing multiple omics datasets and associated metadata¹²¹. *Omix*'s core functions operate on this multimodal container, taking it as input and produce processed omics data, quality control metrics, analysis outputs, and a log of selected parameters, all organised within the same object's structured metadata slots. Moreover, analysis results are appended to the object's metadata, enabling users to perform multiple parallel integrative analyses and store the structured integration results within the same object (**Figure 2-1 Multimodal container**). This object-oriented structure, combined with wrapper functions, greatly enhances the user-friendliness of the analysis, reducing it from thousands of lines of code to approximately 10 (i.e. *code implementation in Figure 2-1*, parameters not shown).

2.2.3 Processing and quality control

Each omics layer is independently processed using established protocols, such as DESeq2 for bulk transcriptomics¹²² and Probatch for proteomics¹²³. Users can choose which parameters and steps of the modular sequence they want to execute based on their analysis requirements. The options include imputation, feature filtering, normalisation, transformation, batch correction, denoising, and sample outlier removal. Furthermore, automatic ontology detection enables the translation of features between different ontologies, such as a single gene name from Ensembl ID or UniProt ID, altogether facilitating the interpretation of results.

<i>Steps</i>	<i>Transcriptomics</i>	<i>Proteomics</i>
<i>Feature filtering</i>	Minimum percentage of samples that meet the user-defined minimum gene count.	Proteins with missing values in more than a user-defined percentage of samples are removed; Removes protein outliers: Remove proteins with average protein abundance

		across samples in bottom and upper 5 percentile
Imputation	Not applicable to transcriptomics post pre-processing. Zeros are real observations (i.e. gene not expressed in a tissue). Normalisation and filtering steps are sufficient to handle sparsity.	Remaining missing values are set to 50% of the minimum value to avoid overly skewing the values. The assumption is that the protein is missing because its intensity wasn't sufficient for it to be selected for MSMS in the mass spectrometry (and therefore isn't identified), rather than it being absent altogether.
Normalisation/transformation	Log2 transformation Rlog: Based on median of ratio method (DESeq2) where counts are divided by sample-specific size factors (i.e. accounts for sequencing depth and RNA composition). Variance stabilising transformation (VST): transforms normalised counts to yield homoscedastic counts (having constant variance along the range of mean values). This is useful for subsequent sophisticated modelling including clustering, ML, or linear discriminant analysis.	Available methods for normalisation are Median centering (i.e. shifts the intensity of the sample to the global median of the experiment) and quantile normalisation (i.e. sets different distributions of individual samples to the same quantiles, which forces the distribution of the raw signal intensities to be the same in all samples). These apply on a Log₂ transformed abundance matrix (divides raw abundance by the ratio of sample median to the mean of all the sample medians).
Batch correction & denoising	Uses Limma's <i>RemoveBatchEffect()</i> function: Removes batch effects for specified technical covariates and regresses out unwanted biological covariates such as sex, post mortem interval or brain regions.	Correction for MS-batches can be done using feature-level median centering, and ComBat from the ProBatch package. To further regress out unwanted biological covariates we extract the residuals obtained from fitting a linear mixed model to each protein. However, it strongly relies on the assumption that the relationship with the outcome is linear.
Sample outlier removal	PCA on batch corrected data Z-score > user-defined threshold along PC1 is used to define sample outliers.	
QC	Compares raw data to processed matrix using heatmaps, PCA and distributions. Correlation analyses assess whether correction improved the quality of the data.	
Formatting	Formatting of Ensembl ID to Gene IDs	Formatting of UniprotKB to Genes IDs.

Table 2-1 Processing steps implemented by Omix

2.2.4 Single omic models

Before the integration process, it is imperative to establish a baseline understanding of each omics layer independently, utilising single omic models. *Omix* provides a range of analytical tools, such as differential expression analysis supported by Limma¹²⁴, a conventional method for identifying transcripts/proteins that display differential expression between distinct grouping variables or alongside a continuous outcome. The Weighted Gene Correlation Network Analysis (WGCNA) is also available to identify gene modules associated with specific disease covariates¹²⁵. The Sparse Partial Least Squares (sPLS) method is also offered as an option to define a concise set of omics features, referred to as a “molecular signature”¹²⁶, that can explain the phenotype by identifying the most explanatory variables.

2.2.4.1 Differential expression analysis

Differentially expressed genes/proteins are identified using a linear model of processed gene/protein expression as a function of a user-specified outcome of interest (i.e. pathology level, AD status), adjusted for user-specified disease covariates. The Log₂ fold change is defined as the slope of the regression line. The package uses the *lmFit()* function, followed by adjustment for multiple testing using the Benjamini & Hochberg (BH) method. Features with an adjusted p-value below 0.05 are considered statistically significant. Significant genes and proteins from the two platforms are compared by plotting Log₂FoldChanges from the transcriptome versus Log₂FoldChanges from the proteome. They are categorised as either *concordant* (i.e. matched direction of effect for a single feature at the gene and protein levels) or *discordant* (i.e. unmatched direction of effect for a single feature at the gene and protein levels). The remaining features are either not significant in both platforms or uniquely differentially expressed in the transcriptome or the proteome.

2.2.4.2 Sparse Partial Least Squares (sPLS)

Sparse Partial Least Squares (sPLS) is a supervised dimensionality reduction method that identifies a subset of features from high-dimensional data that are most predictive of an outcome variable, while also accounting for correlations among predictors. sPLS models are tuned and run separately for each omic modality, where tuning is performed internally via the *tune.spls()* function from mixOmics on a user-provided range of features to retain to train the

model. The number of selected features is chosen based on the lowest mean squared error (MSE) across a range of tested parameters, while the number of components is selected using the Q^2 criterion, a cross-validated measure of model predictive ability that reflects how well the components generalize to unseen data¹²⁶. A univariate sPLS regression model is built using tuned parameters, one layer of processed omics data as a predictor, and an AD continuous trait of interest (i.e. global pathology, microglial activation load) as the response variable. The model seeks linear combinations of a subset of predictor variables that are maximally covariant with the response variable. The final omic signature is defined as the combination of genes/proteins that have either positive or negative loadings depending on the sign of the response variable loading (i.e. if it is positive, then predictor features with positive loadings contribute to the response variable; if it is negative then predictor features with negative loadings contribute to the response variable).

While other regularization methods such as Lasso or Elastic Net are not currently implemented in *Omix*, they could be readily integrated to enable alternative or complementary feature selection strategies.

2.2.4.3 WGCNA

The WGCNA algorithm available from the WGCNA R package (version 1.71) is used for the gene/protein network analysis, as previously described in Langfelder et al. (2008)¹²⁵. The approach involves calculating the pairwise correlation between each gene or protein pair to obtain a signed adjacency matrix. Then, the adjacency is turned into a Topological Overlap Matrix (TOM), and hierarchical clustering of the genes/proteins is performed on the dissimilarity matrix (1-TOM). Network modules can be detected using a “dynamic tree cut” approach. Modules can be functionally annotated via GO enrichment analysis. Module eigengenes/proteins, equivalent to the first PC of the expression matrix of the genes/proteins in each module, are used to identify disease modules and their gene/protein drivers, highly correlated to clinical traits of interest.

2.2.5 Multi-omics vertical integration

Omix's model selection was driven by the need to balance performance, flexibility across omics layers (i.e. with future package updates supporting additional omics beyond transcriptomics/proteomics), and interpretability. Additionally, I ensured to select models from packages that are well maintained, preferable over more complex or less widely adopted alternatives. *Omix* offers a range of state-of-the-art, validated algorithms for vertical multi-omics integration designed to accommodate the three most common integrative research scenarios¹¹³:

- 1. Exploration of molecular mechanisms:** Multi-Omics Factor Analysis (MOFA) and its temporal extension MEFISTO provide a lower-dimensional data representation by inferring a common latent space, enabling correlations between factors and biologically relevant covariates¹⁰⁶.
- 2. Sample stratification:** iCluster is an effective tool for integrative clustering of multi-omics data¹²⁷.
- 3. Biomarker discovery:** DIABLO and its continuous counterpart, sparse multi-block PLS (sMB-PLS) from mixOmics, facilitate variable selection and molecular signature identification¹⁰⁵.

Omix employs the custom *integrative_results()* function to return standardised integrative results across different algorithms presented in sections 2.2.5.1 and 2.2.5.2. This returns a multi-omics molecular signature using a custom strategy tailored to the specific model employed. In MOFA, the signature comprises features from a factor strongly correlated with the user-specific clinical variables of interest and weights above or below 1.5 SD from the mean (default). The signature is the selected sparse set of features for sparse multiblock PLS. In the case of multi-omics clustering, the signature consists of differentially expressed features between clusters. Ultimately, this signature serves as input for downstream analyses, including multi-omics networks, module analysis, and functional, cell-type, and transcription factor-target enrichment tests described in section 2.2.6.

2.2.5.1 Unsupervised multi-omics integration via matrix factorisation

2.2.5.1.1 MOFA

The mathematical foundations of MOFA have been previously described in detail in Argelaguet et al. (2018)¹⁰⁶. Briefly, MOFA is an unsupervised method designed to decompose multiple omics datasets into a shared latent factor space. Each omics dataset Y^i is represented by an integrated factor matrix Z , and each omic layer is associated with a weight matrix W^i that captures the contribution of each feature to the latent factors. MOFA allows for separating shared and omic-specific variability, with residual noise modelled as omic-dependent. This approach helps uncover underlying biological signals across different data types while accounting for noise and variation specific to each dataset.

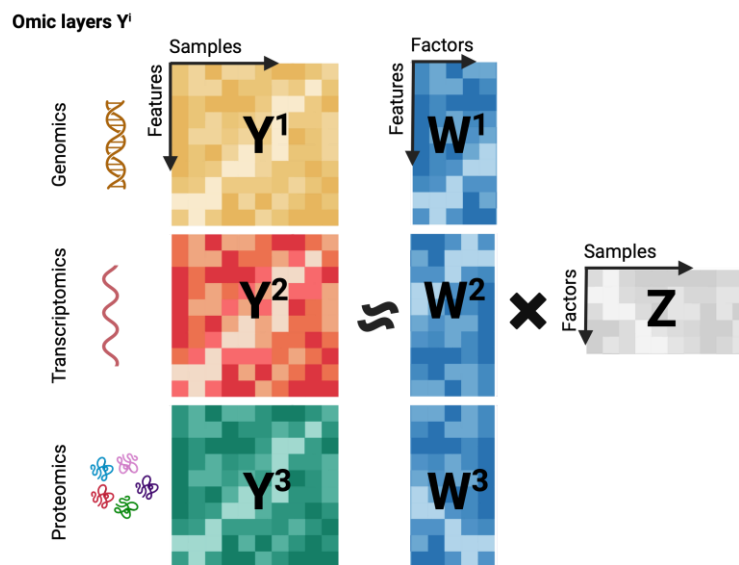


Figure 2-3 Schematic view of MOFA model

Figure generated via [Biorender](#). Adapted from Akalin et al. (2020)¹²⁸

2.2.5.1.2 iCluster

The mathematical foundations of iCluster have been previously described in detail in Mo et al. (2018)¹²⁹. Briefly, iCluster takes a Bayesian approach to latent variable modelling, where it infers distributions over model parameters rather than estimating a single maximum-likelihood value. Each omics dataset X^i is modelled as a linear combination of a coefficient matrix W^i , a shared latent variable matrix Z , and Gaussian noise ϵ^i . A Gaussian prior is imposed on Z , and the likelihood is optimised through a log-likelihood function. To avoid overfitting due to the high number of parameters, iCluster applies Lasso regularisation, penalising the absolute

values of parameters to encourage sparsity and improve interpretability. The Expectation-Maximization algorithm estimates the latent variables and parameters until convergence iteratively.

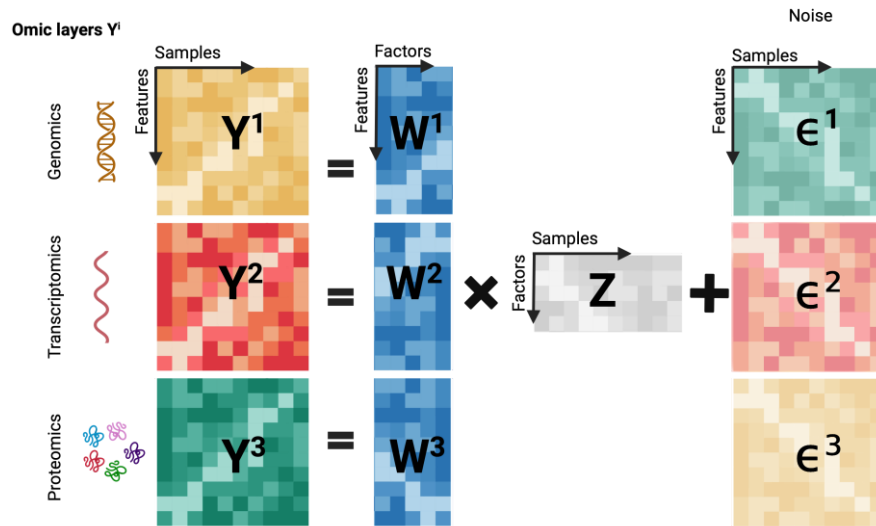


Figure 2-4 Schematic view of iCluster model
Figure generated via [Biorender](#). Adapted from Akalin et al. (2020)¹²⁸

2.2.5.2 Multi-omics integration via supervised PLS

PLS (Partial Least Squares) maximises the covariance between datasets by identifying components that reduce the data's dimensionality, while sparse PLS (sPLS) applies lasso penalisation to the loading vectors to select the key variables that covary across datasets.

2.2.5.2.1 Multiblock sPLS

The multiblock sPLS's mathematical foundations have been described in detail in Singh et al. (2019)¹⁰⁵. Briefly, multiblock sPLS-DA or DIABLO (Data Integration Analysis for Biomarker discovery using Latent components) is a supervised multi-omics variant of Partial Least Squares Discriminant Analysis (PLS-DA) designed to integrate multiple omics datasets and maximise covariance between latent components across different omics layers. It is built on Generalized Canonical Correlation Analysis (GCCA)¹³⁰. It allows dimension reduction by identifying linear combinations of omics layers X^i that maximise the covariance between datasets. Lasso regularisation is applied to each omics layer X^i to perform variable selection, identifying the most discriminant features in each dataset. This multivariate method identifies a correlated and discriminant multi-omics signature, comprehensively integrating multi-omics

data while allowing for feature selection across layers. Comparably to the weights in MOFA, this works by defining a set of loading vectors made of weights for each feature and indicating its relative importance in explaining the response variable (case-control).

While DIABLO focuses on maximising covariance between omics datasets for classification tasks, sMB-PLS is used to model the relationships between different omics blocks by maximising the covariance between the latent components across datasets without the need for predefined class labels. sMB-PLS is typically used in regression tasks, where the goal is to predict continuous outcomes or model correlations across multiple omics datasets rather than performing discrete classifications.

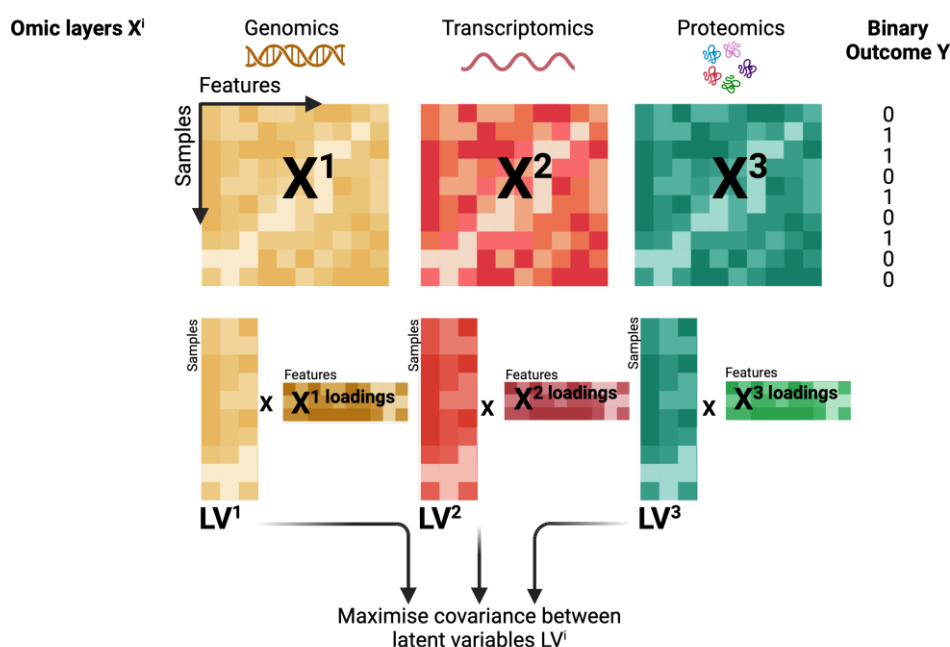


Figure 2-5 Schematic view of the DIABLO model

Figure generated via [Biorender](#). Adapted from Le C  o et al. (2021)¹³¹

2.2.6 Joint Transcriptomics-Proteomics post-integration downstream analyses

The joint Transcriptomics-Proteomics post-integration downstream analyses are developed to extract significant biological interpretations post-integration and implemented into a series of functions overview in **Table 2-2**. Going beyond identifying molecular signatures, *Omix* implements co-expression networks to extract co-expressed features between transcriptomic

and proteomic layers also associated with the clinical phenotype of interest. From the multi-omics signature derived in the integration step, pairwise correlations between features are assessed, and a multi-omics network is built where feature pairs exhibit a correlation exceeding a user-defined threshold. The module analysis then furthers this network, dissecting it into functionally relevant modules through community detection algorithms like Louvain. These essentially build “disease-relevant modules” that have practical implications for understanding disease mechanisms. Module eigenvalues are computed via Principal Component Analysis (PCA) of scaled transcriptomics and proteomics expression and assigned to the 1st PC¹³². Modules are then functionally characterised via pathway enrichment as previously described¹³³, mapped to specific cell types via EWCE¹³⁴ and evaluated in pseudotime analyses with trajectories derived from the Slingshot algorithm¹³⁵. Finally, transcription factor enrichment analysis within *Omix* is a powerful tool for uncovering the regulatory layers of the network. It focuses on identifying transcription factor (TF) proteins whose target genes are present within specific disease modules. This method leverages an external dataset of TF and their target genes generated from previous ChIP-seq experiment data⁷⁴. A Fisher’s exact test assesses whether the overlap between TF target genes and each disease module is statistically significant. Through this approach, *Omix* also sheds light on regulatory network changes in disease state. *Omix* also allows users to query the [OpenTargets](#), an open platform that integrates data from various sources to prioritise therapeutic targets¹³⁶, providing a reference point for validating results against previously identified targets (i.e. can serve as ground truth). Altogether, the range of downstream analyses available is the core strength and differentiator of *Omix* compared to other multi-omics software packages (**Table 2-3**).

<i>Downstream analysis module</i>	<i>Omix Functions</i>	<i>Goal</i>
<i>Co-expression networks</i>	<i>multiomics_network()</i>	Visualise and map the co-expression links between omic features from the multi-omic disease signature. Represents the pairwise correlation between omics features using Pearson’s correlation.
<i>Disease module analysis</i>	<i>communities_network()</i> <i>multiomics_modules()</i>	Identifies modules using chosen clustering method. Available methods include edge_betweenness, leading_eigen, walktrap, fastgreedy or Louvain available from the igraph package ¹³⁷ . Computes the module eigen value (PC1 of scaled transcriptomics/proteomics expression) and correlates it to chosen covariates.

Celltype enrichment	<i>cell_type_enrichment()</i>	Expression weighted cell-type enrichment (EWCE ¹³⁴) computes and enrichment p-value using random permutations to test if a given gene list has specific expression in each cell type. A snRNA-seq dataset including 76,533 nuclei derived from two <i>post mortem</i> human brain is used as background (Allen brain atlas ¹³⁸).
Pathway enrichment	<i>pathway_analysis_enrichment()</i>	Functional enrichment relates candidate genes/proteins to biological functions (i.e. DEG and DEP from single-platform differential analysis or disease modules derived from integration). Pathway analysis is streamlined via an EnrichR wrapper function ¹³³ , and allows the generation of a pathway report using common pathway databases (KEGG, Reactome, BioCarta, MSigDB, Gene Ontology) and finally presenting my internally developed, interactive network visualisations in a HTML report.
TF enrichment	<i>Transcription_Factor_enrichment()</i> <i>circos_TF()</i>	Identifies TF from proteomics and their potential gene targets from transcriptomics data. An external database of TF-target interactions derived from ChIP-seq data ⁷⁴ , is employed to evaluate whether the identified TF from proteomics show significant enrichment for their target genes within disease-specific modules, using Fisher's exact tests. This approach provides insights into potential TF-regulated gene networks that may be driving transcriptional changes within the disease modules.
Target validation	<i>OpenTarget_dataframe()</i> <i>plot_OpenTarget()</i>	Confronts analysis-derived target to current knowledge. The OpenTargets platform ¹³⁶ uses publicly available datasets and scores target-disease associations to assist in drug target identification and prioritisation. The function checks the derived multi-omic signatures by querying the OpenTargets API to gather a <i>Relevance Score</i> based on aggregated information on genetic studies, pathways, drugs, RNA expression and literature text mining.

Table 2-2 Downstream analysis modules implemented in Omix

	<i>Supported omics</i>	<i>Pre-processing</i>	<i>Single omic analysis</i>	<i>Integrative models</i>	<i>Use case</i>	<i>Downstream analyses</i>	<i>Interactive visualisation</i>	<i>Language</i>
Miodin ¹¹⁵	SNP, RNA, methylation, Proteins,	+	-	MOFA	BM	-	-	R
MiBiOmics ¹¹⁶	miRNA, RNA, Proteins	-	-	Co-inertia analysis	BD	Multi-omics networks	+	Web
Muon ¹¹⁷	Single cell omics	-	-	MOFA, WNN	NA	-	-	Python
Movics ¹¹⁸	SNP, RNA, methylation, Proteins	-	-	Range of clustering algorithms	STR	Survival analysis, enrichment	-	R

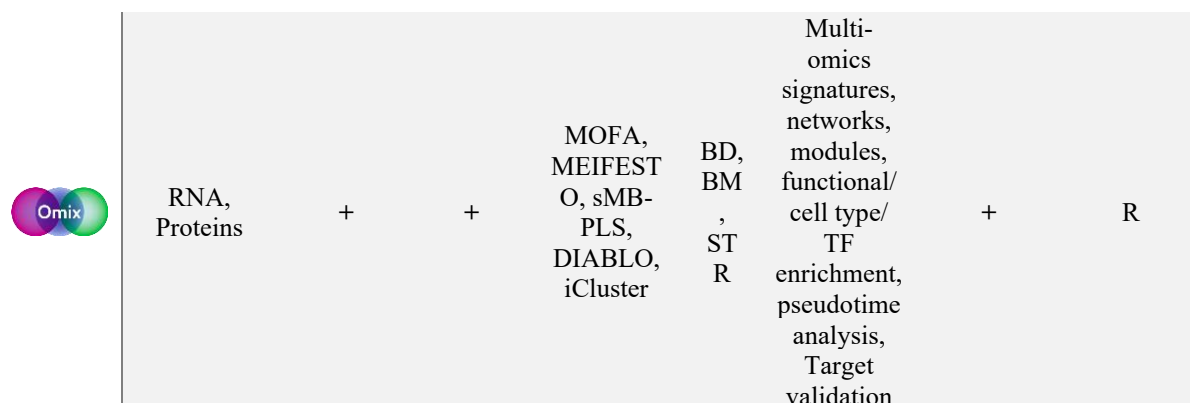


Table 2-3 Comparison of Omix to other multi-omics software packages
Biomarker Discovery (BD), Biological Mechanisms (BM), Sample stratification (STR)

2.2.7 Visualisations and reports

Omix allows users to explore the object's contents directly while also generating interactive HTML reports using its pre-designed templates (**Supplementary figure 2-1:4**). These include a quality control report, a single omic report, a pathway report, and a multi-omics integration report, which display publication-quality plots, tables, and a comprehensive record of the chosen parameters. For reference, full sample reports are accessible on the *Omix* [website](https://eleonore-schneeg.github.io/Omix).

2.2.8 Code and documentation availability

- The source code for the *Omix* package is released under the GPL-3 license and is available at <https://github.com/eleonore-schneeg/Omix>.
- Reference documentation and online tutorials are available at <https://eleonore-schneeg.github.io/Omix>.
- To ensure reproducibility and continuous integration, Github Actions perform build and functionality tests and rebuild package vignettes using a small test dataset after updates. Upon successful completion, a version-tagged Docker image is generated using a Docker file that installs the required software and system-level dependencies. This enables users to conduct end-to-end reproducible analyses without explicitly configuring package dependencies. The Docker image is uploaded to a Docker registry named *Omix* available at <https://hub.docker.com/r/eleonoreschneeg/omix>.

2.3 Methods

2.3.1 Publicly available AD cohort

Brain bulk transcriptomics and TMT proteomics data derived from post-mortem tissue of the dorsolateral prefrontal cortex from individuals enrolled in the Religious Orders Study and Memory and Aging Project (ROSMAP) cohort^{139,140} were used for analysis. Datasets were downloaded from AMP-AD Knowledge Portal (<https://www.synapse.org>) following all terms and conditions to use the data. Complete library preparation and sequencing methods, tissue homogenisation and TMT mass-spectrometry are detailed in ¹⁴¹ and ¹⁴⁰, respectively.

Omix was used to process and analyse ROSMAP transcriptomics and proteomics data. After processing and quality control, the intersection of the 635 individuals with transcriptomics information and the 300 individuals with proteomics information resulted in 205 individuals. The characteristics of the multi-omics cohort are detailed in **Supplementary Table 2-1**. 15,774 protein-coding genes from bulk transcriptomics and 8,165 proteins passed quality control and filtering. Proteins were traced back to the same parent gene, and overall, 7,266 gene equivalents matched in both datasets. The pathological trait selected for analysis is the global pathology score, a summary measure of AD pathology quantifying the burden of A β plaques and NFT over five brain regions. This continuous trait considers all pathological hallmarks of AD and thus provides a more generalisable pathological trait, which can then be narrowed down to specific hallmarks. This approach offers an alternative to categorical grouping, for which a classical Braak staging²⁴ suffered from important group imbalances in this cohort: 40 controls (Braak 0-II), 133 early AD (Braak III-IV), 32 late AD (Braak V-VI).

2.3.2 Evaluation and comparison of transcriptomics-proteomics integration in AD

To investigate whether multi-omics integration can provide a more systematic, accurate and reproducible way to understand molecular mechanisms of AD, I evaluated the performance of two multi-omics integration methods (MOFA, sMB-PLS) against two single-platform approaches (WGCNA, sPLS). I compared molecular signatures derived using these different modelling approaches, as detailed in sections 2.2.4 and 2.2.5. Single omic molecular signatures

included those derived from WGCNA analyses (i.e. modules most correlated to global pathology), sPLS models, both ran separately on transcriptomics and proteomics. Multi-omics signatures included those derived from sMB-PLS and MOFA models. Since these analyses are performed in a discovery setting with no established ground truth, “performance” cannot be assessed using a single criterion. Instead, I considered multiple criteria to evaluate and compare the results of the different modelling approaches:

1. **Biological relevance:** I used the OpenTargets Platform API¹³⁶ to assess and validate the findings' biological relevance, considering current knowledge of omics targets relevant to AD (disease_id = “MONDO_000497” for AD). I retrieved relevance scores from OpenTargets for all available categories: “Known drugs”, “RNA expression”, “affected pathways”, “literature”, “genetic associations” and “animal models”, for each of the feature making up the signature of the different modelling approaches. I computed the mean relevance scores per category and model by summing the relevance scores of all features within each category and model and then dividing by the number of features (1). These mean scores were then summed by model to compute an aggregated relevance score per model (2).

(1)

$$Relevance\ Score_{c,m} = \frac{\sum_{i=1}^{n_{c,m}} Relevance\ Score_{i,c,m}}{n_{c,m}}$$

(2)

$$Aggregated\ Relevance\ Score_m = \sum_c Relevance\ Score_{c,m}$$

Where:

- *i* is a feature (gene/protein), *c* category (i.e. “Known drugs”, “RNA expression”, etc.), and *m* model (i.e. sPLS, WGCNA, MOFA, sMB-PLS).
- *n_{c,m}*: Total number of features within category *c* and model *m*.

2. **Predictive power:** I assessed the performance of the different signatures to predict AD’s global pathology score. To simplify the analysis and mitigate the group imbalance issue, I defined high and low global pathology scores using a median split, where

samples with scores above the median were categorised as 'high' and those below as 'low'. This approach ensures balanced-sized groups needed for a robust binary classification. I trained standard logistic regression models using 80% of the samples, selected through stratified sampling to ensure a balanced representation of high and low global pathology groups, and estimated the performance in predicting a high vs. low global pathology score on the remaining 20% of the dataset. Each model used as input the omic features selected by the signature derived from each single- or multi-omics integration method. Performance metrics, including accuracy, precision, sensitivity, and specificity, were evaluated for each model. While I focused on logistic regression for interpretability and simplicity, alternative methods such as ensemble approaches (i.e. random forests, XGBoost) could be readily implemented to potentially enhance predictive power or capture non-linear relationships between omic features and pathology scores.

- 3. Robustness to down-sampling:** I reran the models on down-sampled datasets at varying levels (10 to 90%) and assessed their level of feature recovery, defined as the percentage of features from the initial signatures that were retained in the results derived from the down-sampled datasets.

2.3.3 Internal data

The internal cohort is part of the UK DRI Multi-Omics Atlas project (map-ad.org). The bulk transcriptomics and proteomics datasets analysed in section 2.4.3 are available in the Synapse repository syn51516099 and were analysed via *Omix* using the processing, integration and downstream analysis steps described in the *Omix pipeline implementation* section. The code and outputs are detailed step-by-step in the *Omix* documentation [vignette](#).

2.3.4 10x Multiome analysis

Data pre-processing including reads alignment and QC was performed by Michael Thomas, bioinformatician in the Matthews' group using an in-house Nextflow pipeline, [Snomix](#). Outputs were passed to Seurat for peak calling and feature selection. Integration was performed using

Weighted Nearest Neighbours (WNN), enabling the joint clustering of cells based on a weighted combination of the two modalities.

I performed the remaining methods for the Multiome analysis:

Combined snRNAseq and ATAC-seq data from microglial cells was analysed using the Signac (v1.12.0)¹⁴² package to identify DNA accessibility changes and TF motif enrichment between AD and control samples. Microglia were kept for differential peak analysis implement via a likelihood ratio test using *FindMarkers(test= 'LR')* and adjusting for sequencing depth (latent.vars = 'nCount_ATAC'). Peaks with increased and decreased DNA accessibility in AD relative to controls were identified. Motif enrichment was conducted with *FindMotifs()* to detect enriched TF motifs, and regions containing STAT1, STAT3, and CEBPB motifs were identified using *GetMotifData()*. Differential peaks were linked to genes within the ME5 microglial module using *LinkPeaks()* with a distance of 1e+06 base pairs, generating peak-TF-gene triplets (“Regulons”) to reveal regulatory relationships. This step pairs peaks with genes where accessibility significantly correlate with gene expression.

2.4 Results

2.4.1 Partial concordance between transcriptomics and proteomics

The differential expression analysis of transcriptomics data yielded 609 and 533 up and downregulated genes with greater global pathology, respectively (absolute Log₂FoldChange >0.1 and adjusted p-value < 0.05) (**Figure 2-6 a**). For proteomics data, there were fewer proteins with a significant Log₂FoldChange over 0.1 (p-value < 0.05), namely 58 in the positive direction against 12 in the negative direction (**Figure 2-6 b**). Since the magnitude of changes is not directly comparable between platforms, I assessed whether significant differential features overlapped between the two platforms and, if so, whether their direction of effects was concordant or discordant (**Figure 2-6 c**). I found most features to be significant only in one platform, with 390 features significantly differentially expressed at both the mRNA and proteomics levels (p-value < 0.05, Log₂FoldChange >0). I found an R=0.32 Pearson’s correlation between pairwise Log₂FoldChanges (p-value <0.05), indicating a moderate positive correlation between the mRNA and protein level changes, reinforced by discordant pairs. Pathway analyses revealed differences between concordant and discordant pairs, whereby

discordant pairs with up-regulated proteins fell into lysosomal and endocytosis pathways, while concordant pairs with up-regulated proteins and transcripts were enriched in cell migration pathways and protein phosphorylation.

The WGCNA approach was applied separately to transcriptomic and proteomic datasets to identify modules of co-expressed genes and proteins associated with global pathology. For each modality, the soft-thresholding power was selected using the *pickSoftThreshold()* function to approximate scale-free topology, with a power of 8 optimised for transcriptomics, and 4 for proteomics. Other parameters were kept to default. Module detection parameters, including minimum module size and module merging thresholds, were selected to balance resolution and biological interpretability. This analysis identified 76 gene modules and 55 protein modules some of which were strongly correlated to global pathology (***Supplementary figure 2-5***). Six protein modules were correlated to global pathology (absolute Pearson's coefficient $R > 0.2$), such as M43, enriched for extracellular proteins and showing the highest overall correlation to global pathology ($R=0.69$, p-value <0.05). Meanwhile, gene modules' correlation with global pathology ranged between $R=-0.13$ to $R=0.2$, highlighting a low relative correlation between gene modules and phenotypes explored here.

2.4.2 Multi-omics signatures provide complementary insights into AD mechanisms

I compared molecular signatures of AD pathology, that is the set of genes and proteins derived from single (WGCNA, sPLS) and multi-platform models (sMB-PLS, MOFA) (**Figure 2-7 a**). Pairwise Fisher's exact tests between molecular signatures revealed significant overlap of features between all methods (p-value <0.05) (**Figure 2-7 b**), suggesting that while the methods identify distinct subsets of features, they also converge on certain key genes and proteins that are likely central to the pathology of AD. VGF emerged as the single marker common to all approaches, highlighting its potential importance in AD pathology. sMB-PLS and sPLS displayed the highest strength of association (OR=78.5, p-value = 2.5e-36), followed by the two multi-omics signatures, sMB-PLS and MOFA (OR= 45.8, p-value =2.5e-35). This could be explained by the fact that multi-omics methods can identify markers of AD pathology identified from a single-platform method while also identifying novel markers missed by single-platform approaches. This is consistent with functional enrichment results, where 5/13 of the top enriched terms derived from multi-omics methods overlap with at least one single-platform approach (i.e. cell migration and synaptic-related terms). In contrast, other terms, including the metal ions and cell migration pathways, are multi-omics specific pathways (**Figure 2-7 c**). **Supplementary figure 2-6** further dissects this finding to show pathway features and the direction of effect, as upregulated or downregulated with increasing AD pathology. While 12/30 terms overlap between single and multi-omics approaches, some, including metal ions pathways in the positive direction and Golgi-related terms in the negative direction, were only found in multi-omics approaches. Since some pathways also appear to be unique to single-platform findings, single and multi-omic approach provide thus provide widely complementary results.

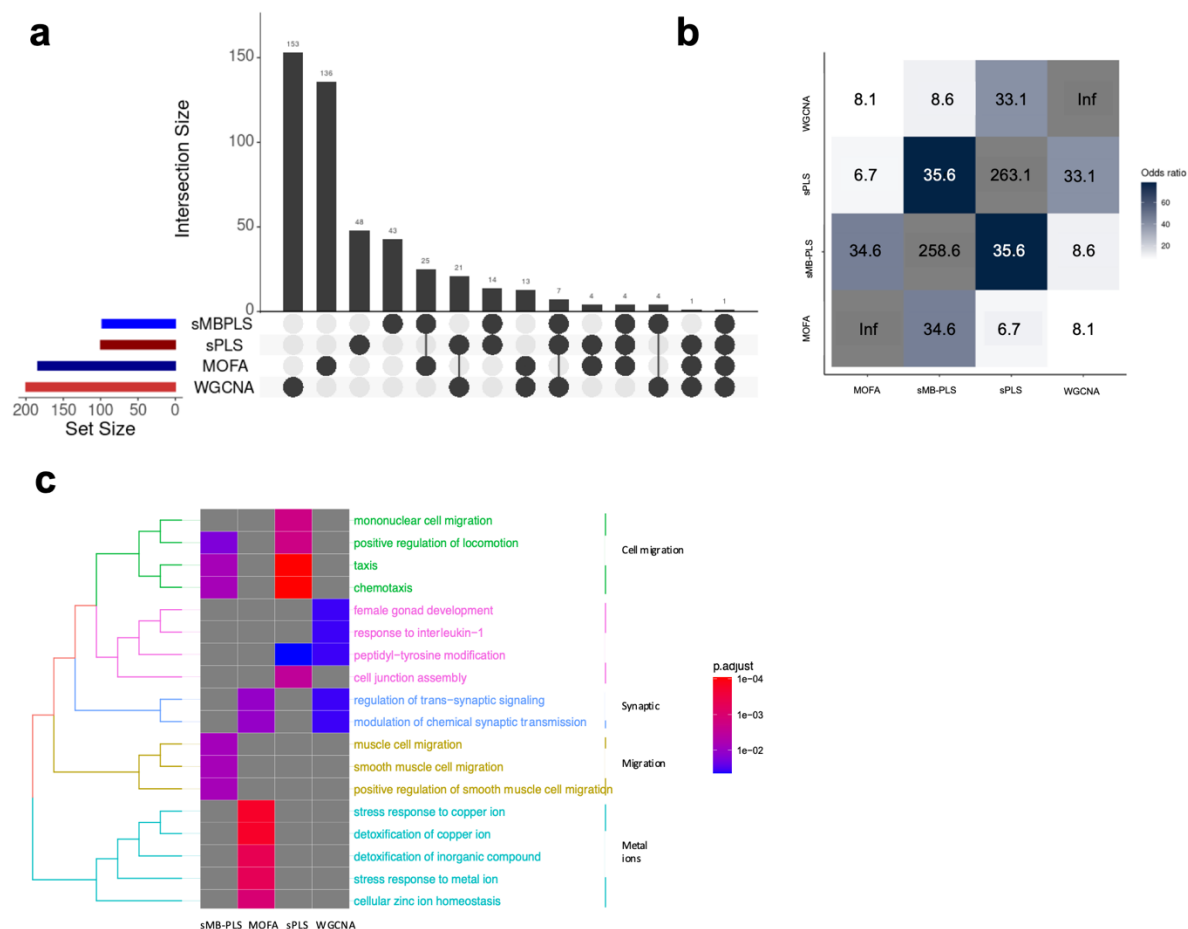


Figure 2-7 Multi-omics approaches partially overlap with single platform signatures
a) Upset plot comparing overlap of signatures derived from each approach, namely from multi-omics models in purple (MOFA and sparse multiblock PLS) and single platform models in blue (sPLS and WGCNA). **b)** Heatmap of Fisher's Exact test between molecular signatures derived from each approach. Labels indicate the $-\log_{10} p$ -values of the association tests. The odds ratio is derived from Fisher's Exact test using pairs of signatures and the union of genes represented by both platforms as background. **c)** The terms are clustered based on similarity and are the top enriched pathways for each approach. Proteomics and transcriptomics signatures from single platform models have been grouped to ease comparison.

Since one cannot exclude the possibility that new pathways may arise from false positive features, I assessed each signature's validity based on biological relevance, predictive power, and robustness. To evaluate biological relevance, I used the OpenTargets platform, a comprehensive resource that integrates evidence from genetics, omics, and literature to score gene-disease associations, and extracted AD relevance scores for each feature (**Figure 2-8 a**). Overall, the two multi-omics approaches consistently produced high aggregated relevance scores, demonstrating strong biological validity. Among the single-omics models, only the sPLS model reached a comparable level of relevance, where its aggregated mean relevance

scored above 2, similar to sMB-PLS and MOFA (*Figure 2-8 a*). This reflects the fact that the sPLS captures a sparse set of proteins with strong biological relevance.

I then compared the predictive power of each molecular signature in classifying individuals with high versus low global pathology (*Figure 2-8 b*). The WGCNA approach performed the worst, while the sMB-PLS model achieved the highest classification accuracy (0.75) and showed the best sensitivity (0.85) for identifying individuals with high global pathology (*Figure 2-8 b*).

Finally, in the robustness assessment, the sMB-PLS showed the second highest level of robustness to down-sampling, with a gradually increasing precision across the down-sampled experiments (*Figure 2-8 c*). Single-platform models based on genes showed variable recovery levels from one experiment to the next, except sPLS on proteomics, which showed the best overall recovery rate.

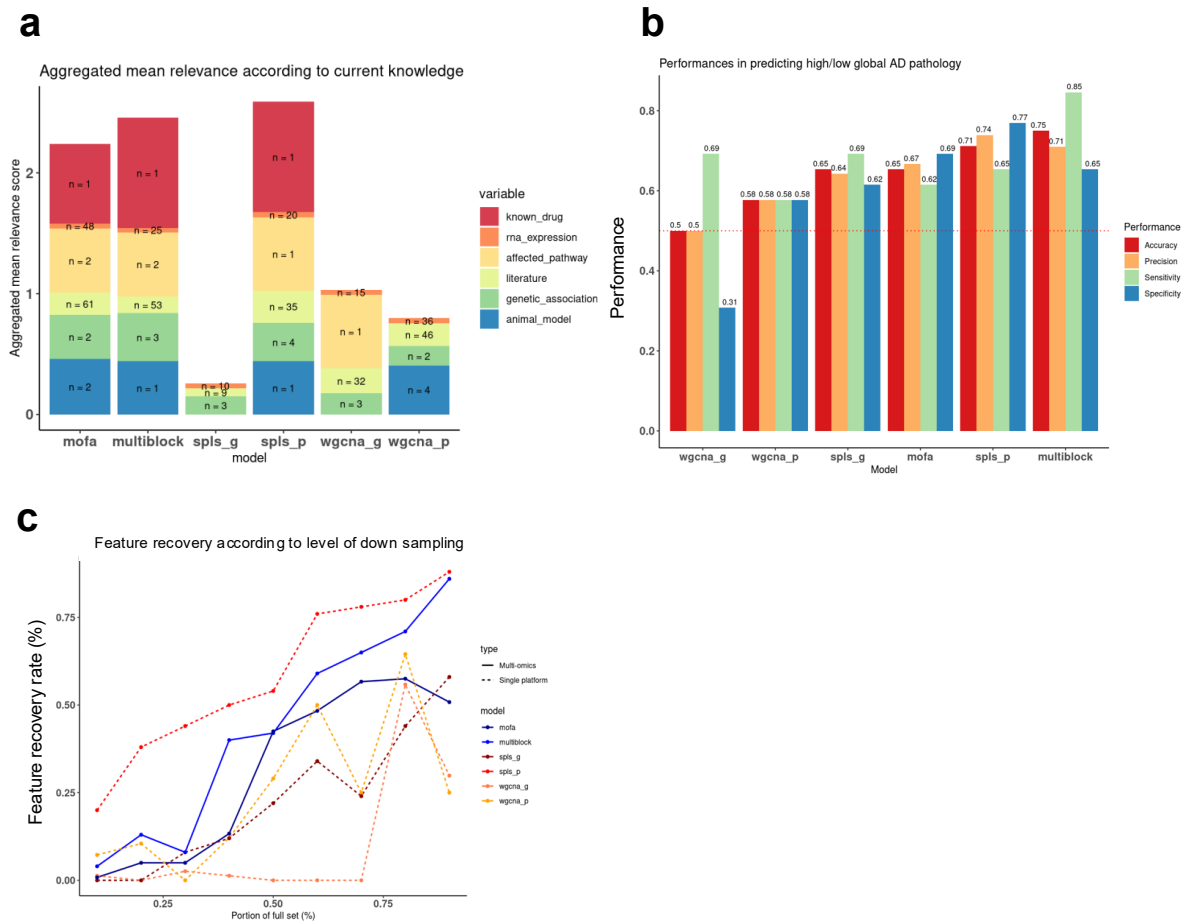


Figure 2-8 Multi-omics approaches are biologically relevant, predictive and robust

a) Aggregated mean relevance score for each model, where ‘_p’ and ‘_g’ denote single omic models on proteins and genes, respectively. n denotes the number of features with available evidence in the database, used to compute the mean relevance score per model per category. OpenTargets derives a 0-1 score based on evidence from an association between a molecular target and a disease of interest, here AD, according to the following categories: Known drugs, RNA expression, affected pathways, literature, genetic associations and animal models. Understudied diseases are unlikely to produce high-scoring targets due to insufficient evidence. In such diseases, a relatively low-scoring target might still be the top-ranked target and potentially a very interesting lead from a therapeutic standpoint. **b)** Performance in predicting high/low pathology group in the test set using logistic regression models. Performance measures include accuracy, precision, sensitivity and specificity. The dotted line at 0.5 indicates the random guess threshold for the classifier. **c)** Signature recovery rate at varying down-sampling levels.

2.4.3 Bulk transcriptomics-proteomics integration uncovers microglia-specific AD regulatory mechanisms

I detail the use of *Omix* in a case study using internal AD bulk proteomics, transcriptomics and 10x Multiome datasets. Specifically, I was interested in characterising a TF-regulated microglia module from bulk data and validating whether the predicted regulatory relationships from the integrated bulk analysis are supported by equivalent changes in snATAC-seq and snRNA-seq in microglia.

40 samples from 26 post-mortem brains ($n=19$ AD, $n=7$ controls) were collected from the mid-temporal gyrus (MTG) and sensorimotor cortex (SOM) regions; the former is known to be affected earlier in the disease based on Braak spatiotemporal staging²⁴ and indeed displayed higher neuropathology load. For most donors (14/26), I used transcriptomics-proteomics samples from both brain regions, while others had only one region available. I performed an unsupervised integration of bulk proteomics and transcriptomics data through the *Omix* pipeline via the MOFA model, to identify shared axes of variation (i.e. factors) relevant to AD pathology (**Figure 2-9 a**). Via *Omix*, I constructed transcriptomics-proteomics co-expression networks to pinpoint disease modules of genes and proteins and identify hub features (**Figure 2-9 b**). I performed differential expression analysis of module eigenvalues between AD and controls to prioritise modules overseeing the largest changes during AD pseudoprogession. Further, I conducted cell-type enrichment analyses using EWCE¹³⁴ against a previously generated cell-type data reference from the Allen Human Brain Atlas¹³⁸ (**Figure 2-9 c**). Factors

1,6,10,15 were considered “AD-specific” factors given these reflect the biological variation attributable to known AD pathological variables such as Braak and phospho-tau immunostainings (AT8, PHF1) (**Figure 2-9 a**). Factor 8 captures variation attributable to brain region differences across samples, displaying correlations with brain regions (**Figure 2-9 a**). From the AD-relevant factors, I derived 30 AD-specific modules (**Figure 2-9 c**), which were further prioritised to 11 based on differential expression analysis (**Figure 2-9 b**). All down-regulated modules corresponded to a neuronal cell type, and up-regulated modules were enriched for endothelial, oligodendrocytes, astrocytes, and microglia (**Figure 2-9 c**). Replication of the 11 prioritised disease modules was sought through the Mount Sinai Cohort (MSBB), an independent large-scale AD cohort with bulk transcriptomics and proteomics data publicly available⁸³. The MSBB transcriptomics-proteomics dataset, including data from 182 individuals, was processed and analysed using *Omix*. The parameters applied were identical to those used in the discovery cohort. I computed replication modules using the same gene and protein features from the discovery cohort modules and found that all but one downregulated module was replicated in MSBB (**Supplementary figure 2-7**).

F6_UP_ME5 (i.e. module 5 derived from positive factor 6) was the most upregulated module in the discovery cohort and was replicated in MSBB. It was enriched for microglia and pathways, including immune response and complement activation. *TREM2* and *HLA-DR*, two well-characterised AD genes¹⁴³, displayed high hub centrality scores in this module, meaning they are the principal intermediaries to other genes or proteins in the module. Another module in the same factor, F6_UP_ME2, was found to be the most upregulated in the MSBB validation cohort. It was found to be enriched for astrocyte lipoprotein metabolism pathways and showcased two important AD genes as network hubs, namely *ABCA1* and *CLU*¹⁴³.

I identified several transcription factors in the proteomics data by overlapping the detected proteins with the ChEA3 transcription factor database, including CEBPB, DRAP1, EEA1, HMGA1, MECP2, STAT1, STAT3, TSC22D1, ZBTB9, and ZNF706. Notably, the microglial module ME5 was enriched for STAT1 and CEBPB target genes and, to a lesser extent, STAT3, though the enrichment for STAT3 was not statistically significant (**Figure 2-9 d,f**). STAT1 and CEBPB showed strong enrichment in microglial targets (p-value = 0.017 for STAT1, p-value = 0.0008 for CEBPB). In contrast, the astrocyte module ME2 was enriched

for MECP2 target genes (p-value=0.044). Further analysis revealed that TF like STAT1, CEBPB, and MECP2 had negative weights in the MOFA analysis, and given the negative correlation of factor 6 and AD variables, a higher protein expression of STAT1 and CEBPB is likely associated with later stage of AD pathology (**Figure 2-9 e**).

I attempted to validate these findings using 10x Multiome data (snRNA-seq and snATAC-seq) from the same internal cohort. The differential peak analysis in AD vs. controls focused on microglial cells (n=1821, **Supplementary figure 2-8**). Then, differential peaks were assed for transcription factor motifs (STAT1, STAT3, and CEBPB motifs) and linked to genes based on correlation to gene expression, focusing on ME5 microglial genes. A number of differentially reduced accessible peaks (negative Log₂FoldChange) were linked to the ME5 microglia genes, indicating that changes in DNA accessibility at these regions are statistically correlated with the expression of ME5 module genes. Overall, these peaks may function as regulatory elements such as enhancers or promoters (**Figure 2-9 g, Supplementary figure 2-9, Supplementary Table 2-2**). Despite this reduction in accessibility, many of these peaks exhibited negative correlations with gene expression, suggesting that decreased DNA accessibility was associated with increased expression of linked genes. For instance, the top downregulated peak, chr19-35931069:35932252, was linked to the TYROBP gene, showed a -1.8Log₂FC (p-adjusted = 0.0001) but a negative correlation to gene expression. This decreased DNA accessibility appeared specific to microglia in AD donors, relative to other cell types (**Supplementary figure 2-10**). These findings aligned with bulk data showing increased ME5 module activity in AD. Peaks linked to ME5 genes and containing motifs for STAT1, STAT3, or CEBPB were significantly enriched among differential peaks (abs(Log₂FC)>0.5) relative to non-differential peaks (OR = 2.28, p-value = 0.02). However, motif enrichment analysis identified additional candidate transcription factors regulating these peaks, such as NRF1, highlighting the complex regulatory landscape underlying DNA accessibility changes in AD (**Supplementary Table 2-3**).

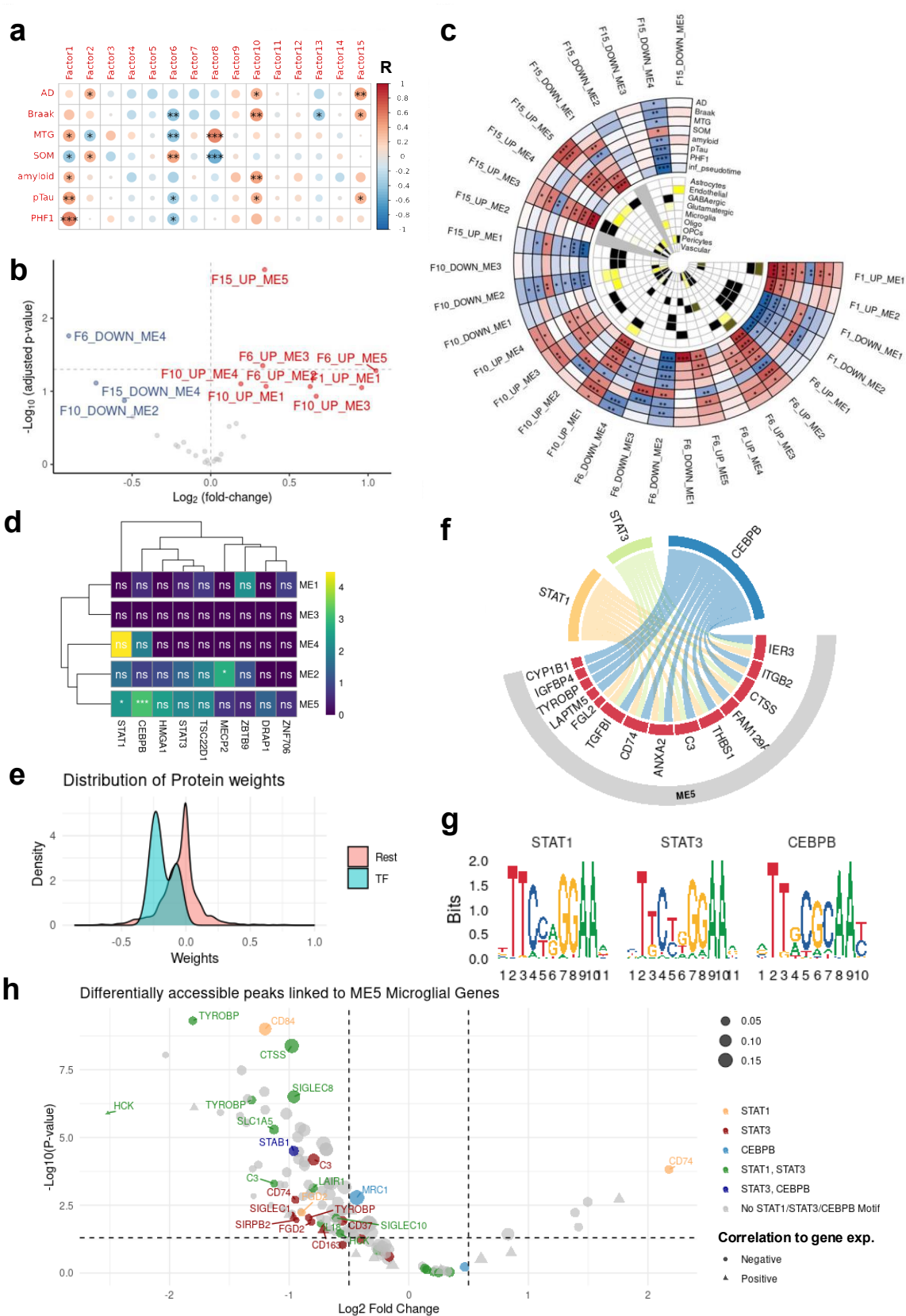


Figure 2-9 Integrated molecular profile of AD highlight complex regulatory mechanisms in microglia

a) Pearson's correlation heatmap between MOFA factors and AD clinical covariates. Significances denoted by * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. **b)** Differentially expressed modules in AD vs. Controls. Models were adjusted for age, sex, and post-mortem interval. In the discovery cohort, modules that were near significance ($0.05 < p_{adj} < 0.15$) were taken for validation alongside significant modules (above the horizontal dashed line). Red shows up-regulated, blue down-regulated modules **c)** Multi-omics transcriptomics-proteomics modules associated with AD. Names correspond to the combination: Factor number 'F#'; Up or Downregulated 'UP/DOWN'; Module number 'ME#'. The outer circle shows correlation measures between module eigenvalues and AD clinical covariates. The inner circle displays EWCE cell type enrichment (black < 0.05 , yellow < 0.15). **d)** Enrichment heatmap of transcription factor targets and Factor 6 modules. Colour shows Odds Ratio and annotation significance level * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. **e)** Distribution of MOFA factor 6 weights for TF proteins STAT1/3, CEBPB, and other proteins. **f)** Circos plot displays enriched TF proteins to target genes in factor 6 ME5 microglia module. **g)** Motif plots of STAT1, STAT3 and CEBPB. **h)** Volcano plot of differentially accessible ATAC peaks linked to ME5 microglial genes in AD vs. control. The x-axis represents the $\log_2$ fold change in accessibility (AD vs. control), and the y-axis shows the $-\log_{10}(p\text{-value})$. Points are coloured based on the presence of STAT1, STAT3, and CEBPB motifs, with shapes indicating whether the peak is positively or negatively correlated with linked gene expression. Peaks with a negative $\log_2$ fold change indicate reduced accessibility in AD. Significant peaks are labelled with their associated, linked genes. Size corresponds to the percentage of cells with accessible peaks in AD.

2.5 Discussion

2.5.1 Complementarity of transcriptomics-proteomics

This study provides the first in-depth approach to characterising AD pathology using integrated transcriptomic and proteomics data. Single platform models highlighted differences between transcriptomic and proteomics changes related to AD pathology, as numerous post-transcriptional mechanisms may be at play.

These single-platform analyses highlighted noticeable differences between transcriptomics and proteomics and should be further studied to differentiate technical versus biological effects. Overall, results highlight that (1) different platforms do not cover the same set of features; (2) of the features that intersect, most see a significant change in one platform only; (3) of the features that significantly change in both platforms, some may see discordant directionality of effects; (4) one platform, proteomics, shows higher relative correlation to pathology.

Many biological processes could explain this uncoupling, such as RNA turn-over/stabilisation or protein translation/degradation, as well as feedback loops. Technical artefacts attributed to data collection may also be involved. Results highlight the higher relevance of proteomics to studying AD pathology, which is consistent with previous studies. Namely, Johnson et al. highlighted AD's proteomic nature using AD-associated co-expression networks at both RNA and protein levels and found only a partial overlap between the two¹⁴⁴. In other words, proteomics captures substantial biological variation related to AD pathology that transcriptomics studies may miss. Since proteomics sits closer to the phenotype than transcriptomics, it may be more directly responsive to the disease state and thus show greater functional relevance.

2.5.2 *Omix* streamlines integration to characterise potential disease mechanisms

Comparisons between single and multi-omics models point to the potential of using transcriptomics-proteomics integration to highlight differences in biological layers and to pick up on features that were not detected with lower statistical power in single omics analyses. In a vertical integration setting, models can capture multiple features' covariance, increasing the sensitivity for identifying molecular features. As such, I could recover previously known molecular markers of AD pathology, including APP, MDK, NTN1 or SPARK¹⁴⁵, and point to understudied markers such as Metallothionein (MT2A, MT1E). Functional pathways related to these markers highlighted novel pathways that may play an important role in AD, in the latter case, of metal ions imbalances, which have been previously hypothesised to be involved in A β production and tau phosphorylation^{146,147}, while others have pointed protective functions against oxidative stress¹⁴⁸. Multi-omics analyses can also re-emphasise findings from single platform analyses, where in this analysis, all approaches pointed to VGF as a key molecular target involved in the pathology of AD. This is consistent with previous studies highlighting the role of VGF as a potential biomarker and therapeutic target in AD^{149,150}. As such, multi-omics signatures displayed the best overall biological relevance when compared to current known targets of AD. On top of improved sensitivity for markers, I showed evidence that a multi-omics approach may be more robust in a context where fewer samples are available and

provide improved predictive ability. Nonetheless, results also highlighted the difficulty of comparative pathways analyses in a discover setting and mechanistic hypotheses ultimately need to be validated downstream with wet-lab experiments.

The case study on internal data demonstrated the value of utilising transcriptomics-proteomics integration to enhance our comprehension of AD molecular processes. *Omix* provided a comprehensive functional profile of all identified disease modules, where hub features (gene or protein) could serve as potential novel therapeutic targets or disease progression biomarkers. This workflow identified several glial mechanisms in microglia and astrocytes, such as complement activation or lipoprotein metabolism, which are known mechanisms of AD pathology progression¹⁵¹. These modules could also be projected into single RNA sequencing analyses to map to cellular states/subpopulation and infer potential protein markers for such states. Furthermore, disease modules could be integrated with additional omics data, such as single nucleotide polymorphisms (SNPs), to investigate how genetic risk factors modulate such disease pathways. This approach provides avenues for modelling causal relationships and deepening our understanding of complex diseases like AD.

2.5.3 Transcriptomics-proteomics integration reveals microglia-specific regulatory mechanisms in AD

Additionally, the integrated analysis of proteomics and gene expression uncovered disease-specific transcription factors, such as CEBPB or STAT1/3, found to be (co-)regulating the gene expression of the ME5 microglia module. This is consistent with previous studies highlighting the role of STAT3 and CEBPB in regulating a neuroinflammatory response in AD^{152,153}. However, while the method implemented in *Omix* provides a broad bulk-level identification of TF and gene networks, it relies on bulk transcriptomic and proteomic data and does not directly give information on cell-type-specific regulatory mechanisms.

Using 10x Multiome data (snATAC-seq + snRNA-seq), I confirmed differential chromatin accessibility of microglia peaks in regions containing motifs for the designated TFs and linked to the expression of target genes of interest (ME5 module genes). One key observation was that chromatin accessibility at STAT1/3 and CEBPB binding sites was reduced in AD microglia, yet the expression of their target genes remained upregulated. Given that previous studies have

identified these TF as activators and not repressors¹⁵⁴, this suggests a complex compensatory mechanism, where increased activity or expression of STAT1/3 might compensate for decreased chromatin accessibility, potentially driving inflammatory gene expression through alternative binding sites or regulatory mechanisms. In other words, in AD, reduced accessibility prevents direct activation by these TFs, potentially causing dysregulation of inflammatory pathways and gene overexpression via parallel compensatory mechanisms. This could be via epigenetic changes involving chromatin remodelling, like histone methylation¹⁵⁵. While open chromatin is generally linked to active transcription, the phenomenon of “epigenomic erosion” reflects a dysregulation where repressive chromatin regions become more accessible, potentially activating genes that are normally silenced, while open chromatin regions become less accessible, impairing the expression of genes essential for normal cellular function. This dual disruption contributes to a loss of regulatory balance in late-stage AD, as observed in a recent epigenetic study by Xiong et al.¹⁵⁶. The main limitation of my two-step approach, which targets TF and gene network candidates from bulk data, was the limited coverage of transcription factors in the bulk proteomics dataset. For instance, SPI1 which was previously proposed to be candidate TF regulator for microglia¹⁵⁶ was missing from this proteomics data. As a result, I focused my search on a restricted set of TFs, which constrained the scope of the single-cell validation.

2.6 Conclusions and future work

I developed *Omix*, an efficient R software package designed to process, integrate, and analyse bulk transcriptomic and proteomic data end-to-end. *Omix* offers a comprehensive range of advanced processing functions, integrative models, and quality control features, enabling researchers to perform diverse integration strategies with speed, scalability, and reproducibility. The tool demonstrated its ability to decipher complex molecular and regulatory mechanisms in AD, by identifying key transcription factors and molecular signatures relevant to AD pathology. This approach highlights the potential of leveraging transcriptomics-proteomics integration in an agnostic fashion to generate novel hypotheses on disease mechanisms and biomarker candidates. At the bioinformatics level, *Omix* provides the foundation for standardising and reproducing multi-omics analyses, emphasising the complementarity between transcriptomics and proteomics for studying disease mechanisms. In the future, *Omix* could be further expanded to support the processing and analysis of additional

omics layers, such as epigenomics and metabolomics. *Omix* is a building block wrapper package used in all subsequent chapters of this thesis to streamline the processing and analysis of transcriptomics and proteomics data.

2.7 Supplementary tables

https://eleonoreschneegans.shinyapps.io/supplementary_2/

Supplementary Table 2-1	ROSMAP test cohort
Supplementary Table 2-2	Peak-TF-Gene table of different peaks in AD vs. Controls
Supplementary Table 2-3	Motif enrichment of differential peaks linked to ME5 microglia genes

2.8 Supplementary figures



Single Omic analyses report

30 June, 2023

Cohort overview

Transcriptomics

Proteomics

Single omic comparison

Single omic report overview

- Cohort overview
- Transcriptomics
- Proteomics
- Single omic comparison
- Functional enrichment

Analysis summary

Analysis name: ADvsControl

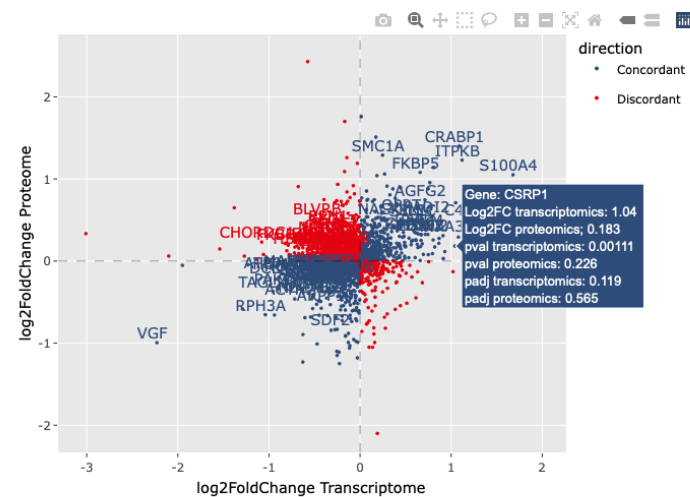
Enrichment database: Reactome_2016

Cohort overview

Table1

	Control (N=20)	AD (N=24)	Overall (N=44)
PHF1			
Mean (SD)	0.0985 (0.236)	1.67 (1.12)	1.20 (1.19)
Median [Min, Max]	0.00623 [0.000519, 0.679]	1.92 [0.00839, 3.57]	0.679 [0.000519, 3.57]
Missing	12 (60.0%)	5 (20.8%)	17 (38.6%)
amyloid			
Mean (SD)	0.422 (0.333)	3.61 (2.68)	2.70 (2.69)
Median [Min, Max]	0.334 [0.0456, 0.968]	3.13 [0.0577, 8.38]	1.48 [0.0456, 8.38]
Missing	12 (60.0%)	4 (16.7%)	16 (36.4%)

Volcano plot



Supplementary figure 2-1 Single omics analyses report



Quality control report

18 April, 2024

Cohort overview

Table 1

Dataframe

Transcriptomics

Proteomics

Multimodal object

Quality control report overview

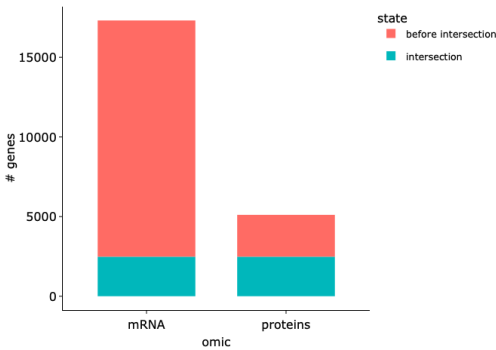
- Cohort overview
- Processing of transcriptomics
- Processing of proteomics
- Multi-omics object

Cohort overview

Table 1

	AD (N=66)	Control (N=46)	Overall (N=112)
age			
Mean (SD)	80.4 (10.8)	80.5 (8.13)	80.4 (9.76)
Median [Min, Max]	83.0 [43.0, 94.0]	80.0 [65.0, 94.0]	80.0 [43.0, 94.0]
sex			
F	34 (51.5%)	24 (52.2%)	58 (51.8%)
M	32 (48.5%)	22 (47.8%)	54 (48.2%)
diagnosis			
AD	66 (100%)	0 (0%)	66 (58.9%)
Control	0 (0%)	46 (100%)	46 (41.1%)
Braak			
3	8 (12.1%)	0 (0%)	8 (7.1%)
4	4 (6.1%)	0 (0%)	4 (3.6%)
5	18 (27.3%)	0 (0%)	18 (16.1%)
5TO6	12 (18.2%)	0 (0%)	12 (10.7%)
6	16 (24.2%)	0 (0%)	16 (14.3%)

Intersection of samples & genes



Correlation of mRNA abundance and protein expression

Dataframe

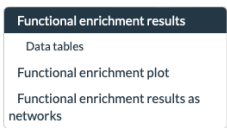
Distribution

Show 10 entries

Search:

corr	gene
-0.776010054671551	PREP
-0.70403019874558	CTD1

Supplementary figure 2-2 Quality control report

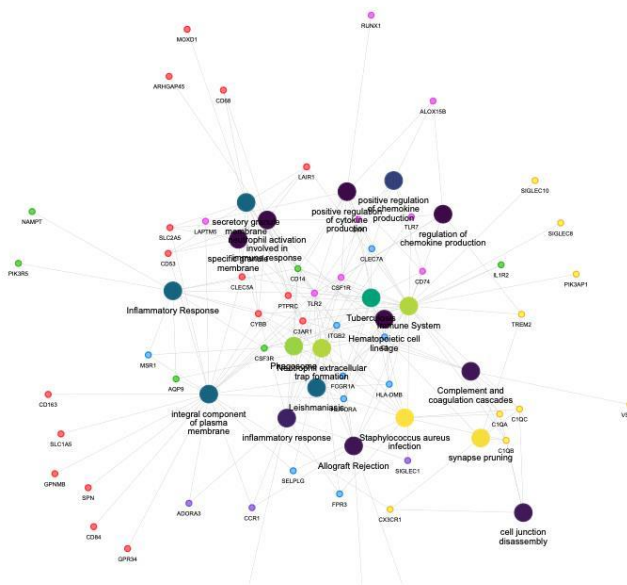


06 December, 2023

- **Functional enrichment**
- **Enrichment database:** GO_Molecular_Function_2021, GO_Cellular_Component_2021, GO_Biological_Process_2021, Reactome 2016, KEGG 2021, Human, MSigDB Hallmark 2020

Data tables

3	MHC class II protein complex binding	17	3	52.01	0.000048	0.00256	CD74;HLA-DMB;HLA-DRA
4	C-C chemokine receptor activity	23	3	36.39	0.00012	0.00408	CCR1;CX3CR1;CXCR4
5	C-C chemokine binding	24	3	34.66	0.00014	0.00408	CCR1;CX3CR1;CXCR4
7	chemokine receptor activity	26	3	31.64	0.00018	0.00408	CX3CR1;CCR1;CXCR4
8	chemokine binding	32	3	25.09	0.00034	0.00668	CX3CR1;CCR1;CXCR4
11	G protein-coupled receptor activity	283	6	5.38	0.0013	0.01889	CX3CR1;P2RY12;GPR34;C3AR1;CXCR4;FPRL1
17	actin binding	177	4	5.64	0.0069	0.06436	SCIN;PARVG;CXCR4;LCP1
19	protein tyrosine kinase activity	108	3	6.9	0.011	0.09186	CSF1R;SYK;SLA



Omix v1.0.0 - 2023-12-06 14:58:27

A report by **Omix**



Multi-omics analysis report

30 June, 2023

Multi-omics integration report overview
Integration overview
Correlation heatmap
Multi-omics signatures
Multi-omics heatmap
Multi-omics network
Multi-omics modules
Cell type enrichment
Functional enrichment
Transcription Factor enrichment
Pseudo-time analysis
Open Targets (Positive weights)
Open Targets (Negative weights)

Multi-omics integration report overview

- Integration overview
- Correlation heatmap
- Multi-omics signatures
- Multi-omics heatmap
- Multi-omics network
- Multi-omics modules
- Cell type enrichment
- Functional enrichment
- Transcription Factor enrichment
- Pseudotime analysis
- Open targets

Integration overview

Analysis summary [Model](#) [Variance explained](#)

Integration model: MOFA
Dependent variable: diagnosis
Sense check variable: PHF1
Clinical covariates: PHF1, amyloid, pTau, AD, Braak
Weights threshold: 0.2
Correlation threshold: 0.4
Disease ID (Open target database): MONDO_0004975

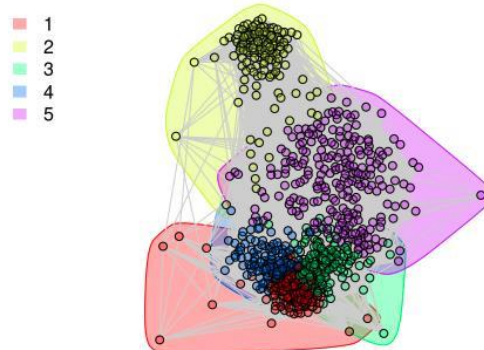
The interpretation depends on the direction of the analysis. If the direction is negative, features with positive weights are 'down regulated' in samples in disease state.

Direction of the analysis: -1

Factor most strongly correlated with sense check variable selected for downstream analysis: 6

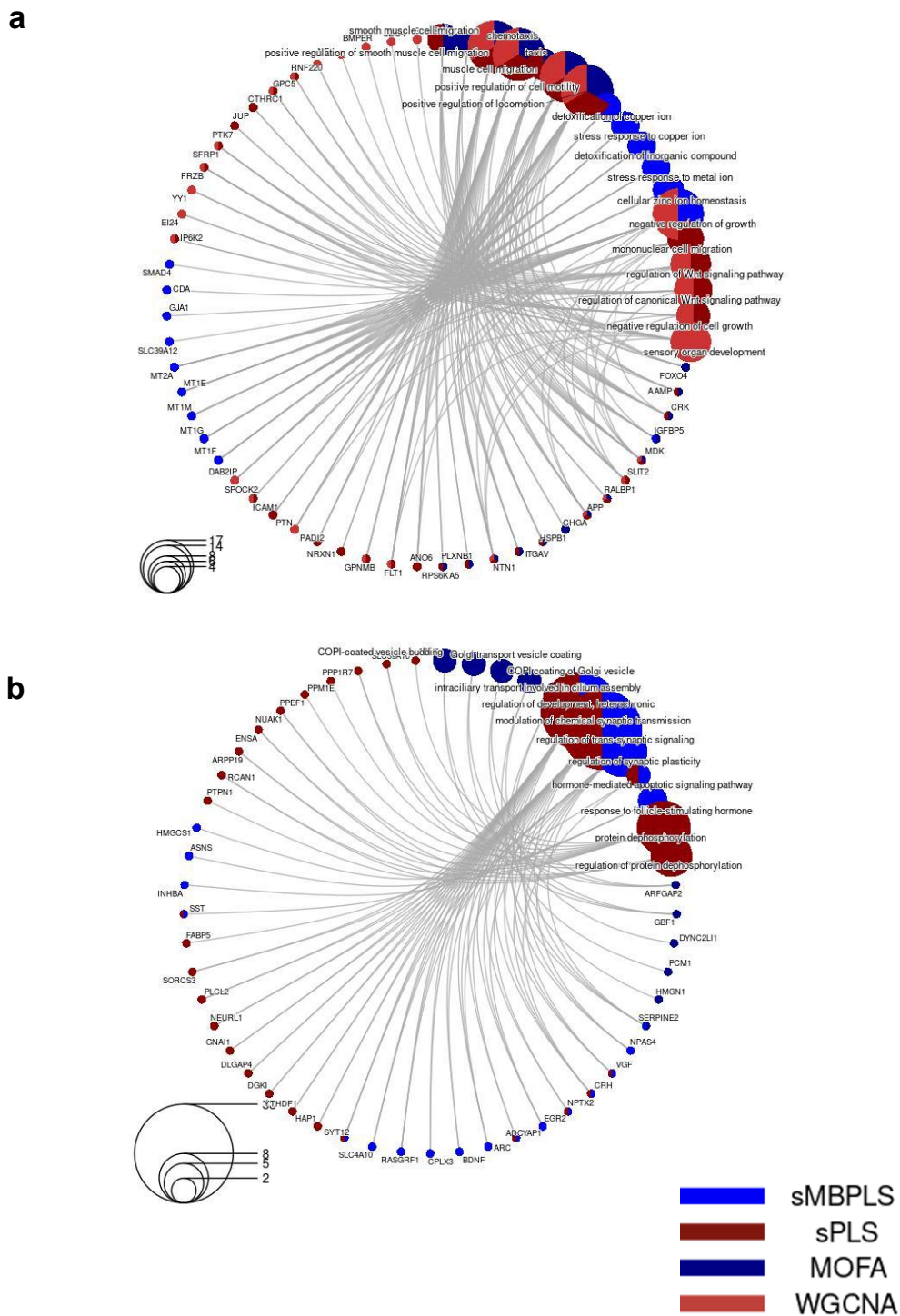
Multi-omics network

[Positive weights](#) [Negative weights](#) [Interactive visualisation](#)



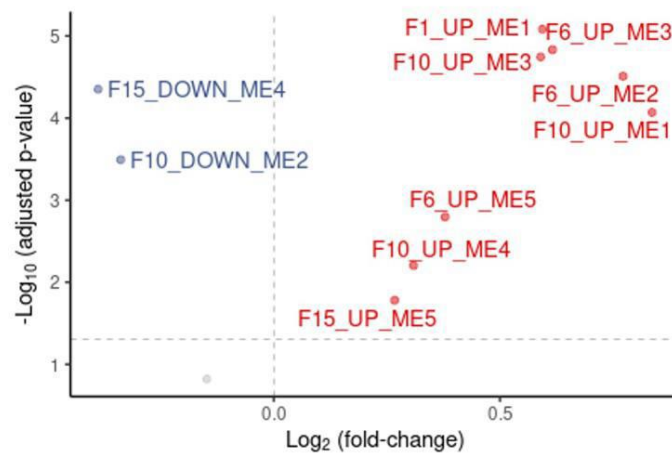
Supplementary figure 2-4 Multi-omics report

amyloid load), and (3) cell-type enrichment via EWCE. (A) The WGCNA analysis on proteins revealed a total of 55 protein modules, of which module 43 (M43) showed the highest correlation to global pathology (0.69, p -value <0.05). M43 was enriched for extracellular matrix proteins such as GPNMB, SMOC1, SPON1, FLT1. Cell-type enrichment revealed a glial enrichment for both astrocytes and oligodendrocytes. M27 was the most strongly negatively correlated with global pathology (-0.3) and enriched for neurons. **b)** 76 gene modules showing a lower relative correlation to global pathology, ranging between $R=-0.13$ and $R=0.2$. The most highly correlated showed a signature enriched for astrocytes, yet no specific enrichment for GO terms (M49). The most strongly negatively correlated (M71) with global pathology showed enrichment for oligodendrocytes and genes involved in mitochondrial processes.

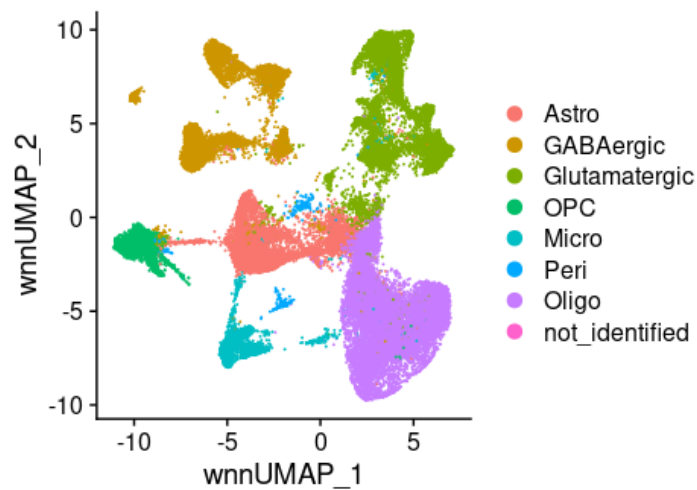


Supplementary figure 2-6 Multi-omics signatures of AD pathology

Network relating proteins/genes to corresponding enriched pathways for the different approaches (dark blue: sMB-PLS, light blue: MOFA, dark red: sPLS, light red: WGCNA). **a)** Up-regulated features **b)** Down-regulated features. The size of the nodes indicates the number of markers from the signature involved in the pathway of interest.

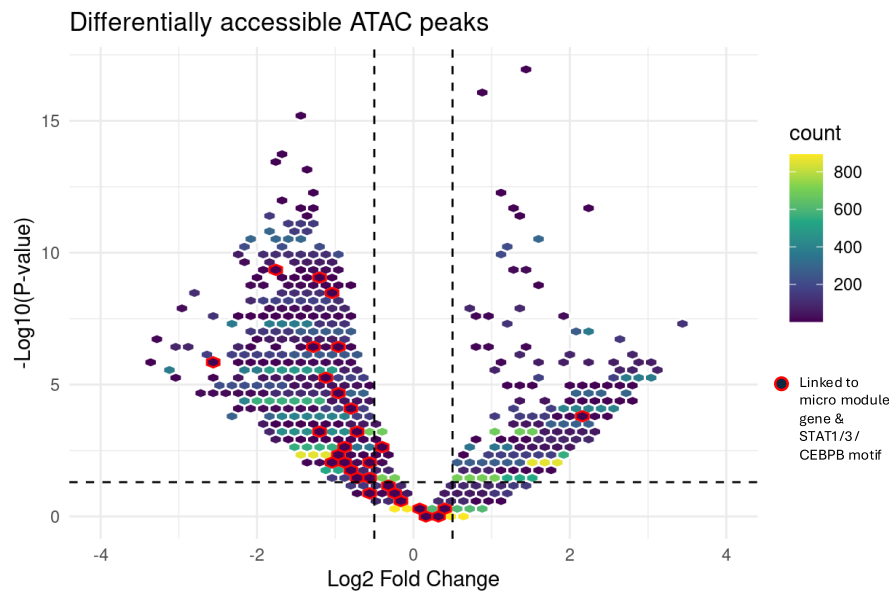


Supplementary figure 2-7 MSBB module replication



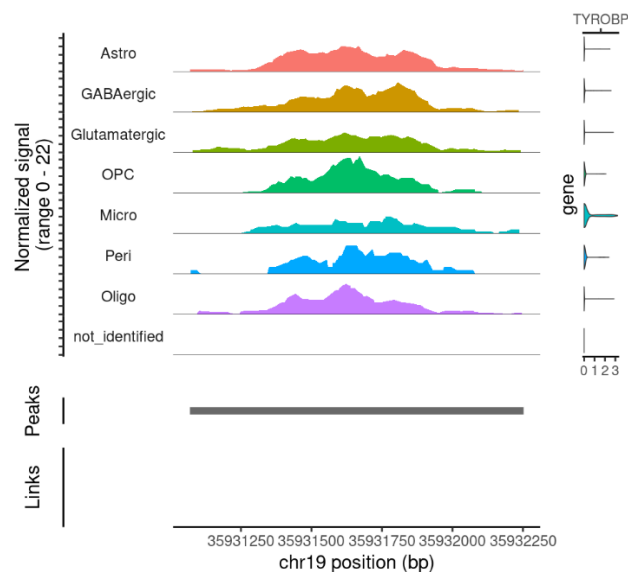
Supplementary figure 2-8 Weighted Nearest Neighbour (WNN) UMAP visualisation of 10x multiome RNA+ATAC from internal cohort

UMAP was constructed using the WNN approach, which integrates RNA expression and chromatin accessibility based on a weighted combination of both modalities. $N=40129$ cells. The microglia cluster is subsetting for analysis.



Supplementary figure 2-9 Differentially accessible peaks

Peaks linked to microglial module ME5 genes and containing motifs for *STAT1*, *STAT3*, or *CEBPB* are highlighted in red. For visual clarity, peaks are binned into hexagonal grids (50 bins), with the colour intensity reflecting the count of peaks in each bin.



Supplementary figure 2-10 Chromatin accessibility in the chr19:35931069-35932252 region and *TYROBP* expression

Regions of higher *Tn5* insertion within the genomic region chr19:35931069-35932252, indicating areas of greater Chromatin accessibility. Data is visualised for AD donors only. Each track corresponds to a specific cell type and has been normalised using a scaling factor that accounts for both the number of cells in the group and the mean sequencing depth of that group. There is an evident reduced Chromatin accessibility in microglia in this region. On the right-hand side, the gene expression levels of the linked gene *TYROBP* are displayed, showing increased expression in microglia.

Chapter 3 Integrative multi-omics analysis of glial signatures associated with accelerated cognitive decline in Alzheimer's disease

Chapter declarations:

1. Some of the work in this chapter has been previously made available as a preprint:

El onore Schneegans, Nurun Fancy, Michael Thomas, Emily Adair, Nanet Willumsen, Marianna Papageorgopoulou, Vicky Chau, To Ka Dorcas Cheung, Robert C. J. Muirhead, Harry Whitwell, Riad Yagoubi, Xiaowen Zhang, Aisling McGarry, Brian M. Schilder, Paul M. Matthews and Johanna S. Jackson (Aug. 2024). “Integrative multi-omics reveal glial signatures associated with accelerated cognitive decline in Alzheimer's disease”. In: *medRxiv*. doi: [10.1101/2024.08.27.24312641](https://doi.org/10.1101/2024.08.27.24312641).

The copyright holder for this preprint is **Eléonore Schneegans**.
[CC-BY-NC-ND 4.0 International license](#).

2. I thank the following contributors:

- M.T Performed bulk transcriptomics pre-processing.

All the internal data used for analysis was generated entirely by the wet lab team, specifically:

- E.A led the IHC and snRNA-seq data generation.
- N.W led the IHC and snRNA-seq data generation.
- M.P snRNA-seq, IHC data generation, and 10x Multiome data generation.
- V.C contributed to IHC data generation.
- D.C contributed to IHC data generation
- R.C.J M contributed to IHC data generation.
- H.W contributed to proteomics data generation.
- R.Y contributed to proteomics data generation.
- A.MG contributed to bulk transcriptomics data generation.
- X.Z contributed to bulk transcriptomics data generation.

Abstract

Background: AD is a neurodegenerative disorder characterised by progressive cognitive decline and memory loss, influenced by both genetic and environmental factors. Synaptic dysfunction and loss are strongly associated with cognitive decline in AD. AD cellular phase plays a key role in the onset of clinical symptoms, as the failure of glial cells to maintain homeostasis may drive the transition from cellular dysfunction to cognitive decline, but the molecular drivers and temporal dynamics of these processes remain unclear.

Objectives: To characterise microglia and astrocyte subpopulations involved in synaptic dysfunction and cognitive decline in AD.

Methods: Using a comprehensive integrative multi-omics approach, I mapped bulk transcriptomics-proteomics modules to single-nuclei transcriptomics, AD polygenic risk (genomics), cognitive trajectories, and cerebrospinal fluid/plasma proteomics across multiple human AD cohorts. I derived AD-specific disease modules from bulk data, focusing on glial activation and synaptic changes. I employed a pseudo-temporal approach to study their association with disease progression and rate of cognitive decline.

Results: Findings demonstrate that activation of microglia and astrocytes occurs in specific cell subsets enriched in AD risk genes. Transcriptomic evidence for early microglial activation precedes immunohistological expression of severe neuropathology. Proteomic markers of astrocytic response appear to be most strongly associated with a molecular subtype of early AD patients who experience accelerated cognitive decline. I also found that brains from donors with a history of more rapid cognitive decline showed evidence for reduced SNAP25-VAMP interactions indicative of synaptic dysfunction, exhibited a higher neurotoxic mixed reactive/exhausted astrocyte response, and were associated with the expression of neuronal markers of injury. Related molecular signatures in cerebrospinal fluid and plasma may provide biomarkers to identify patients at higher risk for rapid cognitive decline

Conclusions: Together, results connect glial activation to synaptic dysfunction and cognitive decline in AD and highlight roles for microglial activation in the genesis of AD and later astrocyte reactivity as a potential determinant of clinical symptom progression. This opens up new avenues of research such that targeting glia, particularly astrocytes, could mitigate cognitive impairment. Additionally, the identified signatures may be used to stratify slow and fast-progressing patients in clinical trials.

Graphical abstract

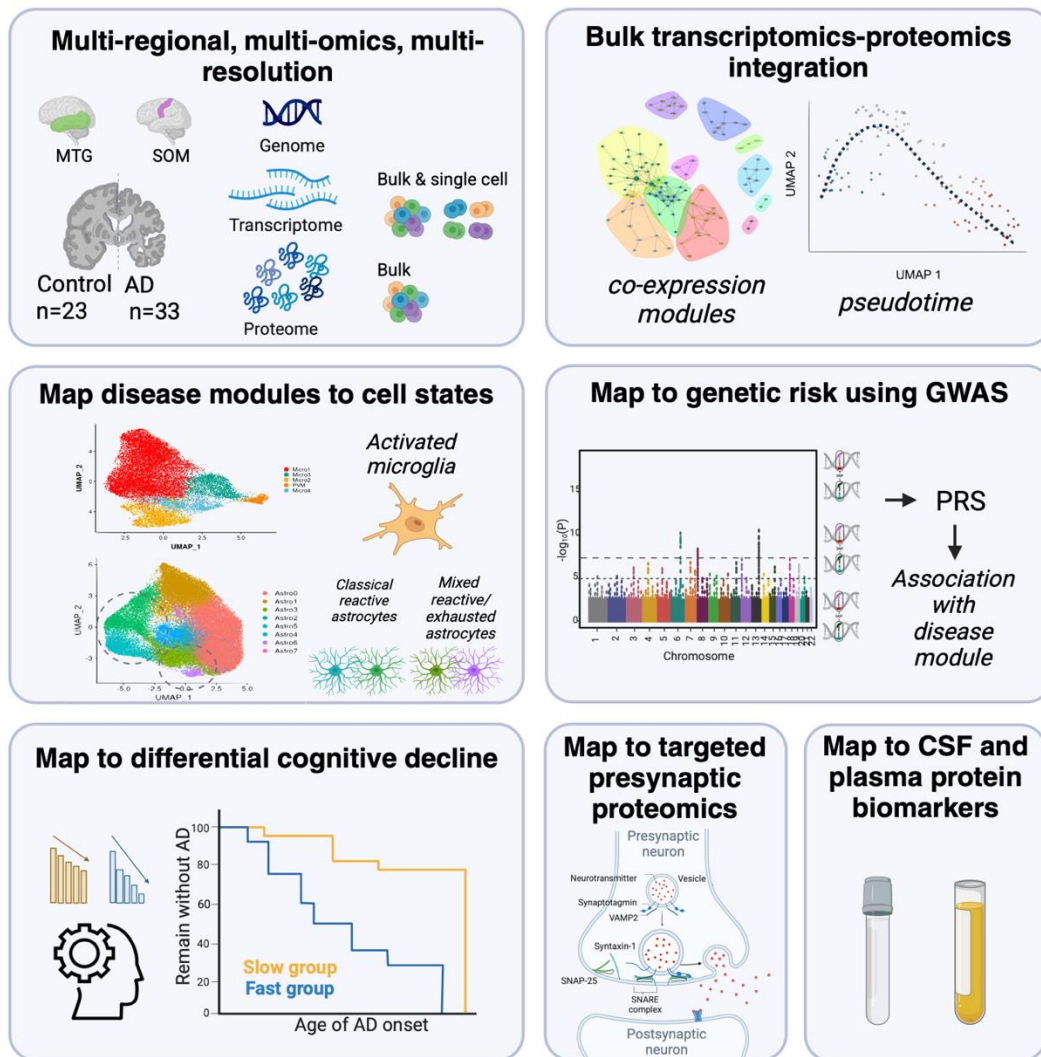


Figure 3-1 A Comprehensive Multi-Regional, Multi-Omic, Multi-Resolution analysis

a) Multi-regional analysis depicting samples from control and AD patients from the middle temporal gyrus (MTG) and somatosensory cortex (SOM), integrating genomic, transcriptomic (both bulk and single nuclei), and bulk proteomics data. **b)** Integration of bulk RNA and proteomics profiling, generation of co-expression modules and derivation of pseudotime. **c)** Projecting bulk modules on single-cell transcriptomics to identify cell state-specific AD modules emphasising activated microglia and classical reactive astrocytes, and mixed reactive/exhausted astrocytes. **d)** Genetic susceptibility to AD was mapped using polygenic risk scores (PRS) derived from GWAS and their interaction with identified AD modules. **e)** Correlation of these modules with differential cognitive impairment. **f)** Targeted SNARE proteomics reflecting synaptic functionality. **g)** Correlation of AD-related modules with CSF and plasma protein biomarkers derived from public datasets. SNARE: Soluble N-ethylmaleimide-sensitive factor Attachment Protein Receptor-mediated exocytosis. Figure generated via [Biorender](#).

3.1 Introduction

3.1.1 Moving beyond amyloid and tau: integrating cognitive data with multi-omics in AD

The neuropathological hallmarks of AD, A β plaques and NFT, trigger a spectrum of molecular disruptions, preceding the neurodegeneration that underpins this disorder. Despite the focus on these pathologies, cognitive decline remains the most clinically relevant endpoint, as it directly correlates with patient quality of life and disease burden^{157,158}. This highlights the need to move beyond a sole focus on amyloid and tau in AD research and calls for integrating cognitive data with omics to uncover the disease mechanisms that precede the onset of cognitive symptoms. With the advent of comprehensive multi-omics cohorts such as ROSMAP¹⁵⁹, MSBB⁸³, and the multi-omics atlas project (MAP), unbiased, untargeted omics approaches facilitate the discovery of molecular signatures across various biological layers, revealing the multifactorial nature and molecular heterogeneity of AD^{65,160}. These cohorts are also well-phenotyped longitudinally during the participants' lifetimes, providing extensive cognitive data that can be linked to post-mortem omics measurements. As such, molecular subtypes of AD have been linked to differential progression trajectories, exhibiting faster or slower progression⁶⁴.

3.1.2 The roles of microglia and astrocytes in AD cellular phase

The cellular phase of AD, marked by glial activation and vascular changes before noticeable cognitive deficits, is now at the forefront of AD research¹⁸. Notably, the study of glial cells' response to neuropathology and their molecular profiling through scRNA-seq and snRNA-seq has yielded substantial advancements in our understanding of AD pathogenesis. Single-cell studies have identified diverse microglial^{29,161} and astrocyte phenotypes^{32,37,40} activated during AD progression. More recent studies suggest that microglial and astrocytic dysfunctions are not mere responses to A β and tau but active participants in disease progression. Namely, Mancuso et al.²⁹ studying microglia, and Serrano-Pozo et al.⁴⁰ or Cameron et al.¹⁶² studying astrocytes, have characterised heterogeneous phenotypes for these glial cells, with different states sometimes exhibiting both neuroprotective and neurotoxic effects. In addition, most GWAS-identified AD risk genes have been shown to converge on microglial response

pathways and influence its response to A β but not to tau pathology, thus highlighting the complex interplay of genetics, pathology, and glial activation in AD¹⁶³. Multi-omics and integrative analyses offer new opportunities for mapping a more comprehensive molecular profile of the cellular changes that occur in AD. Notably, the cellular phase appears central in determining the onset of clinical symptoms, as the failure of glial cells to maintain homeostasis may drive the transition from cellular dysfunction to cognitive decline. Differential glial activation may also partly explain some of the heterogeneity in clinical outcomes, including variations in the rate of disease progression, such as fast or slow progression, as recently suggested⁴⁹.

3.1.3 Aim and objectives

Despite significant progress, the expert consensus outlined in *"Defining Microglial States and Nomenclature: A Roadmap to 2030"*³³ emphasises the need to move beyond the simplistic "good versus bad" perspective on microglial states in disease and instead study these states from a multi-dimensional angle. Similarly, for astrocytes, researchers call for moving away from binary classifications and advocate for more comprehensive, context-specific assessments of their molecular and functional characteristics¹⁶⁴. Glial states are inherently complex and plastic, shaped by the cells' epigenome, transcriptome, and proteome in specific pathological contexts and disease stages. Although the correlation between synaptic loss and cognitive decline is well-established, the role of glial activation — namely, its consequences and temporal sequence relative to synaptopathy and cognitive decline — remains unclear.

In this chapter, I aim to articulate the pseudo-temporal evolution of glial activation and its relationship with synaptopathy and attempt to suggest ways in which glial mechanisms might influence clinical aspects of the disease. I hypothesise that glial activation contributes directly and indirectly to clinical heterogeneity in AD through their implication in neuroinflammation and synaptopathy.

To address this hypothesis, the following objectives have been set:

1. Perform an integrative multi-omics molecular profiling of the AD cellular phase to derive activated glial and neuronal/synaptic dysfunction disease modules.
2. Study how disease modules are enriched and modulated by AD polygenic risk.

3. Integrate molecular findings with longitudinal cognitive trajectories to study the link with the rate of cognitive decline.
4. Map findings to CSF and plasma proteomic data to advance potential candidate biomarkers that identify patients at higher risk for rapid cognitive decline.

3.2 Methods

Sections 3.2.1 to 3.2.5 were not performed and written by me, but are included because they are essential for understanding the data generated and used in this chapter.

3.2.1 Brain tissue

This sub-section of the methods was performed and written by the wet lab team, adapted from methods found in publications from the group^{31,165}.

This study was carried out in accordance with the Regional Ethics Committee and Imperial College Use of Human Tissue guidelines. Cases were selected based first on neuropathological diagnosis (NDC or AD) from UK brain banks (London Neurodegeneration [King's College London], Newcastle Brain Tissue Resource, Manchester Brain Bank, Southwest Dementia Brain Bank [Bristol], and Parkinson's UK [Imperial College London] Brain Bank). MTG and SOM cortex from a final set of 56 donors, including 23 controls (Braak stage 0–II) without evidence of clinically significant brain disease (non-diseased controls, NDC) and 33 AD cases (Braak stage III–VI) were used (total of n=112 brain samples). Cortical samples from two regions were prepared from each brain to characterise pathology, transcript expression and protein abundance.

3.2.2 Bulk transcriptomics

This sub-section of the methods was performed and written by the wet lab team.

Samples were randomized for RNA extraction to reduce batch effects between conditions. Trizol was added to homogenise tissue, followed by Chloroform. Rigorous vortexing of the samples resulted in homogenisation. A Qiagen RNeasy micro kit (cat no 74004) was used for extraction and purification as per the manufacturer's instructions. Briefly, samples in MinElute Spin Columns were treated sequentially with buffers RW1 and RPE. Samples underwent

DNase I digestion, and the resulting RNA was eluted in RNase-free water. Samples were quantified using the nanodrop, and RIN values were obtained using an Agilent Bioanalyser (RNA nano chip and reagents cat no 5067-1512).

Samples were normalised for concentration and underwent ribodepletion using the New England Biolabs rRNA depletion kit (cat no E7400X). Genomic Libraries were constructed using the New England Biolabs Library construction kit (cat no E7760L), using the NEBNext Multiplex oligos for Illumina (cat no E6442S). Libraries were quantified using a Qubit fluorometer and pooled for sequencing using a HiSeq4000 at 50 million reads per sample. Samples were split across multiple lanes to avoid batch effects. The bulk RNA-seq data was pre-processed using the nextflow pipeline nf-core-rnaseq v3.10.1 using default parameters with alignment and quantification performed by the included STAR and RSEM tools. The *post mortem* tissue samples were rejected during quality control if less than 90% of reads could be aligned with a minimum requirement of 30 million aligned reads.

3.2.3 Bulk proteomics and pre-processing

This sub-section of the methods was performed and written by Harry Whitwell, the Proteomics Platform Lead at the National Phenome Centre at Imperial College London.

Tissue samples had been stored in different amounts of optimal-cutting-temperature (OCT) solution. To remove this and standardise tissue preparation, all samples were cryo-sectioned into two 80µm sections, and OCT was removed as previously described¹⁶⁶. Briefly, 1mL of 70% (V/V) ice-cold ethanol was added to the sections with end-over-end rotation for 2 minutes. The ethanol was removed by centrifuging for 2 minutes at 10,000xg at 4°C, and the wash was repeated for a total of 5 washes. Washes were repeated as described, but with water (4 washes) and then 50mM ammonium bicarbonate (3 washes) before allowing the sections to air-dry for 5 minutes. Protein was extracted with 50µL 8M urea in 100mM ammonium bicarbonate with the addition of 0.1mM phenylmethylsulfonyl fluoride (PMSF) protease inhibitor. Lysis was ensured with a probe-sonicator with 10x1second pulses on 30% amplitude, with samples on ice throughout. Lysates were clarified through centrifuging at 10,000xg for 10 minutes, 4°C and the supernatants frozen at -80°C until all protein extraction was completed.

Following a protein assay (DC Protein Assay, BioRad), 20µg of protein was reduced, alkylated and digested using an in-solution digestion approach automated using an Andrews⁺ pipetting robot with a 96-PCR Plate Peltier⁺ (Waters). Briefly, at 4°C, 20µg of protein was combined with water and a pre-mixed digestion solution, resulting in 0.2µg trypsin (Sequencing Grade Modified Trypsin, Promega), 100mM ammonium bicarbonate, 40mM chloroacetamide and 10mM Tris(2-carboxyethyl)phosphine and then heated to 37°C for 16 hours before being chilled to 4°C and acidified to 0.5% (V/V) trifluoro-acetic acid (TFA). Peptides were desalted in a semi-automated manner using an Andrews⁺ pipetting robot in combination with a Vacuum⁺ (Waters) by solid-phase extraction on an Oasis HLB µElution Plate (Waters). Briefly, the plate was pre-equilibrated by washing three times with 100µL 0.1% (V/V) TFA. All samples were added to the plate and washed a further 3 times with 100µL 0.1% (V/V) TFA because being eluted with 50µL followed by 100µL 60% acetonitrile into 96-well QuanRecovery plates (Waters) and dried at 60°C using a Speedyvac (Eppendorf). All volumes were drawn through the plate by negative pressure. All samples were reconstituted into 200µL 0.1% formic acid (FA) by sonicating for 10 minutes in a water bath. A pooled sample was generated from 20µL of all samples and used as a study reference (SR) and for spectral-library generation.

For spectral library generation, 100µg of protein was fractionated into 11 using high-pH reversed phase fractionation kit (Pierce, ThermoFisher) and analysed by LC-MS/MS. Peptide separation was performed using a 0.3x150mm Kinetex XB-C18 column (Phenomenex) in-line with an M-Class HPLC (Waters) using top 60 data-dependent acquisition on a 7600 ZenoTOF (Sciex). Each fraction was injected twice with data acquired over a precursor mass range of 400 to 655m/z or 650 to 905m/z. Precursors were accumulated for 0.1s with MSMS accumulation set to 0.01s and data acquired across 140-1800m/z. Sampled ions were excluded for 6 seconds after a single acquisition with priority given to ions between +2 and +4 charge with greater than 300 counts per second. The peptides were eluted from the LC by increasing the proportion of buffer B (Acetonitrile + 0.1% (V/V) FA) from 3% to 30% compared to buffer A (Water + 0.1% FA) over 44 minutes. B was increased to 80% over the following 2 minutes and held for 2 minutes before returning to 3% for 5 minutes. The flow rate was set to 5µL/minute throughout.

Individual brain samples were analysed by LC-MSMS on the same instrumentation using zenosWATH data-independent acquisition (DIA) with 85 variable windows. MS was acquired between 400-1500 mz with an accumulation time of 0.1s. Over 85 variable windows spanning 399.5 to 903.5 mz , MSMS was acquired between 140-1800 mz using dynamic collision energy ranging from 19 to 43V. Peptides were separated on a 0.3x150mm Kinetex XB-C18 column (Phenomenex) with solvent B increasing from 3 to 30% over 20 minutes before increasing to 80% over 1 minute for a two-minute wash. The column was re-equilibrated to 3% B for 5 minutes before triggering the next sample injection. The flow rate was 5 μ L/minute throughout and 3 μ L of sample was injected in partial-loop mode. The instrument was automatically calibrated every 6 injections, which constitutes 5 samples and 1 SR. All instrumentation was operated through OS version 3.1 (Sciex).

Data-dependent acquisition data was analysed in Fragpipe¹⁶⁷ (version 19) by searching the human SwissProt canonical database downloaded on the 08/09/2022 using defaults. Carbamidomethylation of cysteine was specified as a fixed modification with variable oxidation on methionine. Mass tolerance was +/- 20ppm, and N-terminal methionine was removed. Precursors with a charge between 1 and 4 were allowed with a minimum length of 7 amino acids and a maximum of 2 missed cleavages. Data was exported as a spectral library with FDR specified < 0.01 containing 4636 protein groups and 51105 precursors.

DIA data was searched using DIA-NN¹⁶⁸ (version 1.8.1) using the spectral library described above. Match-between runs was enabled, and DIANN was allowed to determine optimum values for scan width (6), MS (13.6ppm) and MSMS (14.2ppm) tolerance. Outputs were filtered to 0.01 FDR.

3.2.4 Single nuclei transcriptomics

Nurun Fancy, senior bioinformatician, performed the pre-processing, quality control and cell type annotation of the data. Methods are adapted from those found in publications from the group^{31,165}.

The snRNA-seq data was used in a recent publications/preprint from my group, whose samples partly overlapped with the bulk cohort used here (42 out of 112 samples). Single nucleus isolation was performed on 90 tissue blocks from the MTG and SOM regions (n = 27, NDC; n

= 63, AD), following established protocols¹⁴⁸. Approximately 7,000 isolated nuclei per sample were processed and sequenced using the 10x Genomics Chromium Single Cell 3' platform. Pre-processing, quality control, integration, clustering, and cell type assignment of snRNA sequencing data were conducted as described in³¹. The present analysis focuses on astrocyte, microglia, and neuronal subsets. Methods for the sub-clustering and annotation of neuronal subpopulations are detailed in²⁷. Sub-clustering of astrocytes and microglia followed the same steps and used a resolution of 0.1 for astrocytes and 0.2 for microglia, informed by the prior knowledge of expected number of clusters in Mancuso et al.²⁹ for microglia, and Serrano-Pozo et al.⁴⁰ for astrocytes.

3.2.5 Immunohistochemistry (IHC)

This sub-section of the methods was performed and written by the wet lab team, adapted from methods found in publications from the group^{31,165}.

IHC was performed on Formalin-Fixed Paraffin-Embedded (FFPE) sections from the MTG and SOM of each brain studied and paired with material from the cryopreserved contralateral hemisphere used for nuclear preparations for snRNA sequencing. Standard immunostaining procedures, as recommended by the manufacturer, were followed using the ImmPRESS Polymer (Vector Laboratories) and Super Sensitive Polymer-HRP (Biogenex) kits. After dewaxing and rehydration of slides, endogenous peroxidase activity was blocked with 0.3% H₂O₂, followed by antigen retrieval (**Table 3-1**). For immunostaining using ImmPRESS kits, non-specific binding was blocked using 10% normal horse serum. Primary antibodies were incubated overnight at 4°C. Species-specific ImmPRESS or Super Sensitive kits, and DAB were used for antibody visualisation. Tissue was counter-stained by incubation in Mayer's haematoxylin (TCS Biosciences) for 2 min, followed by dehydration, clearing and mounting.

<i>Antibody</i>	<i>Supplier</i>	<i>Dilution</i>	<i>Antigen retrieval</i>	<i>Staining kit</i>
<i>β-Amyloid (4G8)</i>	Biolegend, 800711	1:15,000	Steamer, CB pH6	Super Sensitive Polymer-HRP
<i>pTau (PHF1)</i>	The Feinstein Institutes for Medical Research	1:5000	Steamer, CB pH6	Super Sensitive Polymer-HRP

Table 3-1 Antibodies used for Immunohistochemistry.
CB=citrate buffer.

The remaining methods described in this chapter were designed and carried out by me.

3.2.6 Processing of bulk transcriptomics-proteomics data

Bulk transcriptomics and proteomics data were processed using the Omix R package (v1.0)¹²⁰ to generate a comprehensive multi-assay dataset, with parameters specified in **Supplementary Table 3-2**. After filtering, 20231 genes and 2945 proteins were kept for analysis in the transcriptomics and proteomics data, respectively. 2756 features overlapped between the two modalities. Variations related to sex, *post mortem* delay, and batch effects were regressed out to minimise confounding factors. A quality control (QC) report, including all processing parameters and steps, is provided in **Supplementary File 1**.

External validation of disease modules was sought in two independent large-scale AD cohorts with publicly available bulk transcriptomics and proteomics data (ROSMAP and MSBB). The same parameters described above were used to process these datasets.

3.2.7 Bulk multi-omics integration

Integration of the bulk multi-omics data was performed using the Multi-Omics Factor Analysis (MOFA) model. Prior to integration, scaling of the different modalities was carried out to ensure comparability. Three integration scenarios were considered and explored in the results **section 3.3.1**, namely (1) unbalanced integration, (2) balanced integration on intersecting features, and (3) balanced integration on downsized transcriptomics. This latter was found to be the most appropriate since my primary goal is to identify disease-related molecular features independent of brain region, and the two other integrations could not disentangle brain region-related variation from AD variation. For the integration scenario (3), the 2,945 most variable transcriptomic features (top 14%) were selected to achieve balanced views between the transcriptomics and proteomics modalities. The resulting down-sized transcriptomics dataset had 420 overlapping features with the proteomics. This subsetting was crucial to harmonise the data and enhance the robustness of the integration process. Other integration strategies were deemed suboptimal according to the criteria for disentangling brain-related variation from pathology-related variation (**Supplementary Fig 2.e-f**). Other parameters were kept to default values.

3.2.8 Cross-sectional pseudotime trajectory

To compensate for the lack of longitudinal tissue data in *post mortem* samples, I employed a pseudotime trajectory analysis to model transcriptomic-proteomic dynamics across a disease continuum. It uses the same concept as in pseudotime inference with scRNA-seq¹⁶⁹, except here, each data point is a bulk sample derived from a patient, and pseudotime represents an inferred trajectory along which patient samples are ordered, modelling disease progression from early to advanced stages. To capture molecular changes across different stages of AD, I selected two brain regions, the MTG which is affected earlier in AD and shows higher neuropathological load, and the somatosensory cortex, which is affected later in the disease. The disease trajectory was derived using the slingshot algorithm¹³⁵ on a two-dimensional Uniform Manifold Approximation and Projection (UMAP) representation of AD-relevant and brain region-specific MOFA factors. I clustered the latent space (UMAP1/UMAP2) into 3 groups via k-means clustering and chose the root of the slingshot trajectory as the cluster with a high density of early Braak samples (cluster 2) and ending cluster enriched in later samples (cluster 3). This was specified in the slingshot parameters (start.clus =2, end.clus=3). I then ranked samples according to their position along the slingshot trajectory but within predefined Braak tiers (early [Braak 0-II], mid [Braak III-IV], and late [Braak V-VI]), implementing a semi-supervised approach. The final pseudotime values were determined by fitting a linear model between the derived disease trajectory and the sample ranks within each Braak tier (i.e. fitted values). This provided a continuous measure of disease progression, positioning each patient along an inferred gradient of disease severity.

3.2.9 Transcriptomics-proteomics co-expression modules

The AD-correlated factors 1,7,11 and 14 were selected for co-expression module analysis since they were significantly correlated to both tissue phospho-tau and β -amyloid. Co-expression networks and modules were assessed for each of these factors separately. MOFA features with weights exceeding ± 1.5 SD from the mean were used to define the signature. Pairwise correlations between features in the multi-omics signature were assessed, and a multi-omics network was constructed where feature pairs showed a correlation exceeding a user-defined threshold (default=0.3). This network was then dissected into functionally relevant modules

using the Louvain community detection algorithm¹⁷⁰, resulting in "disease-relevant modules" that aid in understanding disease mechanisms. Module eigenvalues were calculated via PCA on scaled transcriptomics and proteomics expression data, and the first principal component was assigned as the module eigenvalue. Modules identified in the primary cohort were projected onto independent datasets using the same features and eigenvalue computation method.

3.2.10 Differential module analysis

Differential module analysis identified genes with significant expression differences between case and control groups. Linear models using limma (v.3.50.3)¹²⁴ were fitted for each module to assess differential module changes against (1) pseudotime, (2) AD vs. controls, and adjusting for age, sex, PMI, brain region, *TREM2* and *APOE* genotypes, and noise factors (factors not related to AD variation). P-values from the moderated t-tests were adjusted for multiple testing using the Benjamini-Hochberg procedure to control the false discovery rate¹⁷¹. Modules were prioritised based on their significance levels, selecting those with the most significant differential expression for further analysis.

3.2.11 Module cell type enrichment

Cell-type enrichment analyses were conducted using Expression-Weighted Cell Type Enrichment (EWCE) (v1.2)¹³⁴. This analysis utilised a cell-type data reference from the Allen Human Brain Atlas¹³⁸ to determine the enrichment of specific cell types within the identified modules (50277 genes across 9 cell types).

3.2.12 Module functional enrichment

Functional enrichment analysis was performed using Enrichr¹³³. This streamlined pathway analysis allowed for the generation of pathway reports utilising common pathway databases such as KEGG, Reactome, BioCarta, MSigDB, and Gene Ontology. The results were presented using my internally developed, intuitive, and interactive network visualisations¹²⁰ (**Supplementary Files 2-9**).

3.2.13 Projecting bulk modules and gene sets in single nuclei data

To project gene modules identified from bulk RNA sequencing data onto single-cell RNA-seq data, I utilised the *hdWGCNA* package¹⁷². First, I generated single-cell gene expression data in a *Seurat* object containing normalised and scaled expression matrices. I used the *ProjectModules()* function from *hdWGCNA* to calculate module eigengenes in the snRNA-seq data. The projected module scores provide a measure of module activity in individual cells. I subsequently visualised module scores across different cell types or conditions using UMAP, identifying cell types with significant module enrichment. This analysis enables the identification of cellular contexts driving module activity and provides insights into the cellular heterogeneity of gene module regulation.

I used the same approach to project the neuroprotective and neurotoxic gene set on the astrocyte subpopulations. Gene sets were obtained from *Cameron et al.*¹⁶², mapped to human genes and filtered for unique genes to avoid overlapping genes between neurotoxic and neuroprotective genes (**Supplementary Table 3-8**). I also projected the microglia states defined in *Mancuso et al.*²⁹ following the same approach.

3.2.14 Microglia and astrocyte cluster differential expression analysis

For the analysis of differential gene expression, I employed the Model-based Analysis of Single-cell Transcriptomics (MAST)¹⁷³ framework to examine the relationship between gene expression and histopathological features (amyloid/PHF1) within microglial and astrocytic subclusters. Each histopathological feature was modelled as a dependent variable in a zero-inflated mixed-effects regression using the formula $\text{zlm}(\sim \text{dependent_variable} + (1|\text{sample}) + \text{age} + \text{trem2} + \text{apoe_group} + \text{sex} + \text{brain_region} + \text{PMI} + \text{pc_mito}, \text{method} = \text{"glmer"}, \text{ebayes} = \text{F})$. The fixed-effect term *pc_mito* adjusted for the relative proportion of mitochondrial gene counts. Models with and without the dependent variable were compared using likelihood ratio tests to identify significant associations. Genes expressed in at least 10% of cells (minimum of two counts per cell) were included in the analysis, with significant differential expression defined as a $\text{Log}_2\text{fold-change} \geq 0.25$ and an adjusted *p*-value < 0.05 . To control for multiple testing, I applied the Benjamini-Hochberg (BH) procedure to adjust *p*-values for the False Discovery Rate (FDR). Results are available in **Supplementary Tables 3-21:24**.

3.2.15 AD GWAS enrichment analysis

To perform genetic enrichment analysis of modules, I utilised MAGMA software (v1.10)¹⁷⁴. First, I extracted SNP locations and p-values from harmonised summary statistics of the different AD GWAS studies, namely from Bellenguez et al. (2022), Jansen et al. (2019) and Wightman et al. (2021)^{143,175,176}. Using the SNP location data and an NCBI gene location file (NCBI38.gene.loc/NCBI37.gene.loc, according to the GWAS), I annotated SNPs to genes with MAGMA's *--annotate* command (annotation window 50kb upstream and downstream). I then used the annotated gene information to conduct a gene-based association analysis using reference genotypes from the 1000 Genomes Project (European population). The gene-based p-values were computed using the summary statistics file and the previously generated gene annotation file. To identify significant genetic associations across multiple studies, I performed a meta-analysis using gene-based association results from the three GWAS studies (Bellenguez_magma.genes.out, Janssen_magma.genes.out, and Wightman_magma.genes.out) and combined results using the *--meta* command to obtain meta-analysed p-values (meta_analysis.genes.out) (**Supplementary Table 3-9**). For the genetic enrichment analysis of astrocyte and microglia modules, I used the individual and meta-analysis gene-level statistics with MAGMA's *--set-annot* command on a text file containing module entrez-id features. Genes found to be driving the genetic enrichment (p-value<0.05) and present in modules are summarised in **Supplementary Tables 3-11:12** for the microglia (results derived from the meta-analysis) and astrocyte (results derived from the Wightman's GWAS) module, respectively.

3.2.16 Transcription factors target gene enrichment analysis

Transcription factor enrichment analysis serves as a powerful tool to uncover the regulatory layers of the network. It focuses on identifying transcription factor (TF) proteins whose target genes are present within specific disease modules. This method leverages an external dataset of TF and their targets, generated from previous ChIP-seq experiment data⁷⁴, and assesses whether the overlap between TF target genes and each disease module is statistically significant through Fisher's exact test.

3.2.17 Microglia and astrocyte module specific PRS

To compute polygenic risk scores (PRS), I used the PLINK software (v1.90)¹⁷⁷ based on whole genome sequencing data from the ROSMAP cohort and GWAS summary statistics. The SNPs included in the scores were selected from those mapped to genes identified as significant ($p < 0.05$) in driving genetic enrichment within microglia and astrocyte functional modules, as determined in the AD GWAS enrichment step (n SNP for microglia score = 2682, n SNP for astrocyte score = 4174). To calculate the PRS, I applied PLINK's scoring function to derive separate scores for astrocytes (calculated using effect sizes and p-values from the Wightman et al. GWAS) and microglia (calculated using effect sizes and p-values derived from Jansen et al. GWAS). Scores were computed across a range of predefined p-value thresholds from 1×10^{-7} to 0.5, after which both the astrocyte and microglia PRS were scaled to ensure comparability.

3.2.18 Rate of cognitive decline

To assess the rate of cognitive decline, I computed participant-specific rates over a 3-year interval. The data were grouped by participants, and the rate of cognitive decline was measured by calculating the difference in cognitive scores between the most recent time point (i.e., the time of death) and the closest available follow-up year to 3 years before death. This individual rate measures the speed of cognitive decline over time, where a negative rate indicates a decline in cognition, while a positive rate indicates an improvement. I restricted the time window to be close to the time of death to homogenise the varying follow-up timing in the ROSMAP cohort and to ensure that these rates of cognitive decline accurately reflect brain changes quantified *post mortem*. Of the 132 participants, 126 had available longitudinal measurements, with at least 3 measurements per participant. For 90.5% (114/126) of the participants, these measurements spanned a true 3-year interval. In a few cases, measurements were taken at intervals closest to the 3-year span, specifically 2-year (6.35%, 8/126), 4-year (2.38%, 3/126), and 5-year (0.794%, 1/126) intervals.

3.2.19 Multi-omics subtyping of early AD

I employed *iClusterBayes()* from the iClusterPlus R package (v1.22.0)¹²⁷ to cluster AD participants. First, I optimised the number of clusters within the early AD ROSMAP subset

(n=132) using the Cluster Prediction Index (CPI) from the IntNMF package (v1.2)¹⁷⁸ and the Gap-statistics from the MOGSA package (v1.28)¹⁷⁹. Data pre-processing included transforming the values of each dataset to be non-negative by rescaling them to a 0-1 range in order to meet the non-negativity criteria of IntNMF. Using the *nmf.opt.k()* function from the IntNMF package, I calculated the CPI by performing non-negative matrix factorisation (NMF) across a range of cluster numbers, providing a score for each number of clusters tested. Concurrently, the Gap-statistics method, implemented via the *moGap()* function from the MoCluster package, compared the within-cluster dispersion against a null reference distribution, producing a Gap statistic for each number of clusters. The optimal number of clusters of two was determined by combining the CPI and Gap-statistics results, selecting the number that maximised the sum of the CPI score and Gap statistic. Next, I ran a Bayesian integrative clustering analysis using the *iClusterBayes(type=c("gaussian", "gaussian"))* function on transcriptomics and proteomics to identify common underlying structures. The analysis involved 1000 burn-in iterations and 1200 drawing iterations, with a thinning parameter of three to reduce autocorrelation. A prior probability parameter of 0.5, a standard deviation of 0.5, and a beta variance scaling factor of 1 were used. The posterior probability cutoff for cluster membership assignment was set at 0.5 (default).

3.2.20 Differential molecular subtype analysis and pathway enrichment

Differential molecular subtype analysis was performed using the *clustering_DE_analysis()* function from the Omix package. Covariates included PMI, age, and sex. A Log₂fold change threshold of 0.1 was applied for proteomics and 0.5 for transcriptomics. Comparisons were made between cluster sets CS2 and CS1, representing fast versus slow groups, respectively. Multiple testing was adjusted using the BH procedure. For pathway enrichment analysis, Enrichr was used to generate pathway reports. The analysis included several pathway databases: GO Molecular Function 2021, GO Cellular Component 2021, GO Biological Process 2021, Reactome 2016, KEGG 2021 Human, MSigDB Hallmark 2020, and BioCarta 2016.

3.2.21 Kaplan-Meier analysis

I used the Kaplan-Meier method to analyse the delay in clinical AD onset among participants classified into two cluster groups. The analysis was conducted using the *survfit()* function from the survival package (v3.2-3) in R. Participants were considered event-free until the clinical onset of AD (Clinical cognitive diagnosis summary, $\text{dcfdx} \geq 4$ [*Clinical cognitive diagnosis summary*]). The follow-up period was capped at 12 years, based on the findings that longer follow-up times skewed survival estimates and thereby caused survivor bias (n= 104/132, **Supplementary figure 3-4**). Individuals who never experienced AD onset were censored at the time of their last follow-up (death). I compared the survival curves of the two cluster groups, focusing on the time to AD onset, with curves visualised using the *ggsurvplot()* function.

3.2.22 Random forest model to predict pseudotime

The internal dataset was partitioned into training (80%) and testing (20%) using the *CreateDataPartition()* function (caret package in R [v.6.0.94]) to ensure that the distribution of the outcome (pseudotime) is preserved in both the training and testing sets. I then trained a regression-based random forest model with the 15 significant transcriptomics-proteomics modules as predictors and pseudotime as the target variable. Hyperparameter tuning was performed using a grid search over mtry (1-15) and ntree (1000-2500) values within a 10-fold cross-validation, with Root Mean Squared Error (RMSE) as the performance metric. The optimal hyperparameters (mtry = 3, ntree = 1500) were identified and used to train the final model on the entire training dataset (**Supplementary figure 3-5**). The final model performance was evaluated on the test set, yielding an RMSE of 1.105, R^2 of 0.655, and MAE of 0.931. Variable importance was assessed using two metrics: the percent increase in mean squared error (%IncMSE) and the increase in node purity (IncNodePurity), providing insight into the key biological processes underlying pseudotime variation. The trained random forest model was applied to predict pseudotime on the external early AD ROSMAP dataset using the same projected modules initially identified in the internal dataset.

3.2.23 SNARE proteomics

Quantitative immunoassay studies are all performed in the Honer Lab at the University of British Columbia (Vancouver BC, Canada). Detailed ELISA methodologies were published earlier¹⁸⁰. Briefly, this was done using direct ELISA to measure levels of presynaptic proteins implicated in SNARE-mediated exocytosis (syntaxin-1, SNAP-25, VAMP) and capture ELISA to quantify SNARE protein-protein interactions. 76 out of the 132 early ROSMAP participants (n=40/70 slow and n=36/62 fast) have measurements for synaptic density (immunodensities of SNARE proteins) and synaptic protein-protein interactions. This latter is proposed to be an indirect measure of synaptic functionality by Honer et al.¹⁸⁰, as efficient SNARE protein interactions are essential for vesicle fusion and neurotransmitter release, and their disruption leads to impaired synaptic communication.

3.2.24 Linear mixed-effects model with interaction terms

Interaction Models for Cognitive Trajectories (Supplementary Table 3-14):

I employed a linear mixed-effects model using the *lmer()* function from the lme4 package (v.1.1)¹⁸¹ to evaluate the interaction between age and clusters on cognitive trajectories. The dependent variables were global cognition scores (cogn_global, cogn_ep, cogn_ps, cogn_po, cogn_se, cogn_wo). Fixed effects in the model included interactions between age (year) and clusters (cluster_group), with covariates such as sex (msex), baseline age (age_bl), baseline cognition (cog_bl), and years of education (educ). A random intercept for each participant (projid) accounted for intra-individual variability. The age variable was derived by summing baseline age and follow-up year. An age threshold of 80.74 was identified at the cognitive level intersection between the two groups. Interaction effects were visualised using the *interact_plot()* function from the interactions package.

(Equation 3-1)

*Cognitive domain [global/ episodic/ perceptual speed/ visuospatial/semantic] ~ age*cluster group + sex + age baseline + cognition baseline + years of education + (1|projid)*

Interaction Models for a mechanism of synaptic functionality (Supplementary Tables 3-18:20):

I used interaction models focusing on the SNAP25-VAMP synaptic complex (synap_3cort_capture4 variable) to study the relationship between synaptic functionality and various factors. I analysed the following models:

(Equation 3-2)

$$\text{Rate of decline} \sim \text{SNAP25-VAMP Synaptic functionality} * \text{cluster group} + \text{sex} + \text{age death} + \text{years of education} + \text{global pathology}$$

This model examines how the rate of cognitive decline preceding death relates to SNAP25-VAMP synaptic functionality and its interaction with cluster group, controlling for sex, age at death, years of education, and global pathology (gpath).

(Equation 3-3)

$$\text{SNAP25-VAMP Synaptic functionality} \sim \text{astro} * \text{cluster group} + \text{sex} + \text{age death} + \text{years of education} + \text{global pathology}$$

This model looks at the relationship between SNAP25-VAMP synaptic functionality and the astrocyte F14_UP_ME2 module, with an interaction term for cluster group, adjusting for the same covariates.

(Equation 3-4)

$$\text{SNAP25-VAMP Synaptic functionality} \sim \text{micro} * \text{cluster group} + \text{sex} + \text{age death} + \text{years of education} + \text{global pathology}$$

This model investigates the relationship between SNAP25-VAMP synaptic functionality and the microglia F14_UP_ME1 module, including an interaction term for cluster group and controlling for the same covariates.

3.2.25 CSF and plasma proteomics modules overlap

CSF and plasma proteomics modules and relevant annotations were taken from a previous analysis by Dammer et al.¹⁸². I used Dammer's modules, which showed significant enrichment of neuronal, astrocyte, or microglia cell types and were significantly differentially expressed

in AD versus controls. The overlap between my modules and CSF and plasma proteomics modules was tested using Fisher's exact test via the Gene Overlap package (v4.4). This test calculates a p-value to evaluate whether the observed overlap is likely if the gene sets are independent. It also provides an odds ratio, a measure of the association strength between the gene sets. I considered all expressed genes as the background set, following default settings.

3.2.26 Cognitive resilience classification

Resilience can be defined in various ways, but I adopted the framework from Mathys et al.¹⁸³, where the direction of the difference between observed and expected values of cognitive level Y determines resilience and vulnerability. The baseline "expected" value is predicted from another variable, Z, here neuropathology, associated with cognition (Y). The relationship between the observed-expected score (i.e. residuals) and a variable X, which can be a biological process, here, a neurotoxic astrocyte score, tells us how X may contribute to the individual's deviation from expected cognition, thereby shedding light on potential mechanisms of resilience or vulnerability in the disease process. Positive residuals can be interpreted as resilience (better than expected Y), whereas negative residuals can be interpreted as vulnerability (worse than expected Y). A negative relationship between the residuals and the neurotoxic astrocyte score suggests that higher astrocyte scores are associated with more negative residuals, meaning individuals with higher neurotoxic astrocyte activity tend to perform worse than expected on the cognitive or synaptic outcomes (Y) relative to their neuropathology (Z).

Following this approach, I assessed cognitive resilience (CR) in a cohort of 296 individuals, comprising 149 AD patients and 147 controls, using CSF proteomics, ELISA quantifications of tau and amyloid, and cognitive scores¹⁸⁴. CR status was determined based on three pathological measures: total tau (tTau), phosphorylated tau (pTau), and amyloid-beta 42 (A β 42). A linear regression model was employed to predict cognitive performance, as measured by the Montreal Cognitive Assessment (MoCA), to quantify CR, given the level of pathology. The CR score was computed as the difference between observed cognition and cognition predicted by pathology levels (fitted values on the regression line). Individuals were

classified as “resilient” if they had positive CR scores across all three pathologies (i.e. $CR_{\tau\text{Tau}} > 0$ & $CR_{p\text{Tau}} > 0$ & $CR_{A\beta 42} > 0$); otherwise, they were classified as vulnerable.

Candidate biomarkers were presented in section 3.3.7 “Biomarkers in plasma and CSF: Identifying high-risk groups for rapid cognitive decline”. Any markers from my prioritised modules shown to overlap with features in Dammer et al. differentially expressed neuron/microglia/astrocyte module, regardless of overall module enrichment significance and pairwise matching cell types, were considered in the search space (**Table 3-2**). Subject to availability in the CSF dataset, I selected the top markers representing different cell types and associated to CR to compute the CSF biomarker scores. The microglia score was the sum of C3 and CD14, the astrocyte score was defined by AZGP1, and the neuronal score was the sum of GAP43 and NPY levels. To capture the balance between neuroinflammatory glial activation and neuronal integrity, a composite "neuro-glia" score was calculated by subtracting the combined microglia and astrocyte score from the neuronal score.

<i>Neuronal CSF candidates</i>	<i>Astrocyte CSF candidates</i>	<i>Microglia CSF candidates</i>
VSNL1; GAP43; SNCB; RPH3A; NEFL; KRT17; STMN2; PRKCG; FABP3; NPY; CPLX1; C1QL3; SCRG1	AZGP1; DDAH2; TGFB1; FABP7; GARS1; PLEC	NEB; CD163; CD84; ITGB2; HLA-DRA; VPS26A; VSIG4; CLEC71; C3; CD14; CAPG; GPNMB; LCP1; IL1R2; FCGBP; PTPRC

Table 3-2 CSF candidates

I trained two random forest models to classify individuals into one of four categories: AD-resilient, AD-vulnerable, control-resilient, and control-vulnerable. The predictors used included pTau, tTau, A β 42, age and sex in the baseline model, and the same predictors along the neuro-glia score in the second model. To address class imbalance, Synthetic Minority Over-sampling Technique (SMOTE) was applied during model training, which was performed using 10-fold cross-validation. ROC curves were generated for each class, and the area under the curve (AUC) was computed to evaluate model performance. The models used the caret package (v6.0-94) for model training and validation and pROC (v1.18.5) for ROC curve and AUC calculations. I then tested the performance of the two models (with and without neuro-glia

biomarkers) on an independent dataset of 96 individuals, consisting of 33 AD patients, 31 asymptomatic AD individuals, and 32 controls¹⁸⁴.

3.2.27 Data availability

3.2.27.1 Internal cohort

The internal cohort is part of the UK DRI Multi-Omics Atlas project (map-ad.org).

UK DRI Multi-omics Atlas Project

This study's metadata and patient IDs can be found in the following Synapse repository: [syn61606248](https://www.synapse.org/#!Synapse:syn61606248). The BrainBankNetworkID in the sample sheet must be cross-referenced with the omics datasets to ensure the inclusion of the same donors.

- Proteomics: [syn53424292](https://www.synapse.org/#!Synapse:syn53424292)
- Transcriptomics: [syn52943644](https://www.synapse.org/#!Synapse:syn52943644)
- Histopathology: [syn52658344](https://www.synapse.org/#!Synapse:syn52658344)
- Single nuclei transcriptomics: [syn53165790](https://www.synapse.org/#!Synapse:syn53165790)

3.2.27.2 External cohorts

Selected participants' IDs and related derived features (modules, cluster groups, PRS scores, etc.) used in the current study can be found in the study Synapse repository [syn61606248](https://www.synapse.org/#!Synapse:syn61606248)

ROSMAP

- Transcriptomics and TMT proteomics data: Publicly available at Synapse under IDs [syn4164376](https://www.synapse.org/#!Synapse:syn4164376) and [syn21971722](https://www.synapse.org/#!Synapse:syn21971722), respectively.
- Whole genome sequencing: [syn11707419](https://www.synapse.org/#!Synapse:syn11707419)
- SNARE Synaptic Proteomics: Data can be requested from [RADC Data Requests](https://www.synapse.org/#!Synapse:syn11707419).

MSBB

- Transcriptomics and TMT proteomics data: Publicly available at Synapse under IDs [syn16796121](https://www.synapse.org/#!Synapse:syn16796121) and [syn24995077](https://www.synapse.org/#!Synapse:syn24995077), respectively.

Emory Goizueta Alzheimer's Disease Research Center (ADRC)

- CSF TMT proteomics from Johnson et al. 2020: [syn21441787](https://www.synapse.org/#!Synapse:syn21441787)

GWAS studies

- MAGMA: cncr.nl/research/magma

- 1000 Genomes reference data: Summary statistics are available at cncr.nl/research/summary_statistics

CSF/Plasma Module Memberships from Dammer et al. 2022

- Supplementary information is available at <https://doi.org/10.1186/s13195-022-01113-5>

3.2.28 Code availability

The code for this chapter can be found in the Chapter 3 folder of the Github repository

<https://github.com/eleonore-schneeg/PhD>.

All codes concerning the transcriptomics-proteomics integration and related downstream analyses are available as an R package that can be downloaded from the following GitHub repository: <https://github.com/eleonore-schneeg/Omix>.

The final random forest model to predict pseudotime is available from the following Synapse repository: syn61606248.

3.3 Results

3.3.1 Bulk transcriptomics-proteomics integration

I included bulk transcriptomics and proteomics data derived from 112 samples acquired from the middle temporal gyrus (MTG) and somatosensory cortex (SOM) of 56 *post mortem* brains from neuropathologically diagnosed AD and non-diseased control (NDC) donors of comparable ages (n=33 AD & n=23 NDC, mean (SD) 80.4 (10.8) & 80.5 (8.13) years old respectively, **Supplementary Table 3-1**). The analysis involved an unsupervised integration of bulk transcriptomics and proteomics data through the Multi-Omics Factor Analysis¹⁰⁶ (MOFA) model to identify shared axes of variation (i.e. *factors*) relevant to AD pathology. Given the dimensionality of the proteomics (2945 proteins) compared to transcriptomics (20231 transcripts), I filtered the transcriptomics on the top 14% most variable features (2945/20231) to achieve balanced views as recommended by MOFA developers¹⁰⁶. **Figure 3-2 a-c** displays the different integration strategies tested, including unbalanced integration, balanced integration on intersecting features, and the chosen (optimal) approach described in **panel c**. This approach was found to be the most appropriate since my primary goal is to identify disease-related molecular features independent of brain region, and the two other integrations

could not disentangle brain region-related variation from AD variation. The integration strategy was successful in disentangling brain-related variation from pathology-related variation. The optimised model showed a poor correlation between factors, suggesting a good model fit (**Figure 3-2 d**). The model used K=15 factors, explaining 69% of the variation in the transcriptomics data and 23% in the proteomics data (**Supplementary Table 3-3, Figure 3-2 e**).

Expression of factors 1,7,11, and 14 were correlated with measures of AD neuropathology, β -amyloid [4G8] and pTau [PHF1] (range of Pearson's coefficient R values from 0.26 to 0.46, $p \leq 0.05$, **Figure 3-2 c**). I found significant differences in factor values between early Braak (0-II) and late Braak (V-VI) stages (**Figure 3-2 f**). I found that factor 3 variance was well explained by brain region ($R^2=0.65$, $p \leq 0.001$). This indicates that the MOFA model effectively disentangled AD-related molecular variation from those due to regional differences, as the factors correlated with AD pathology were distinct from the factor associated with brain region.

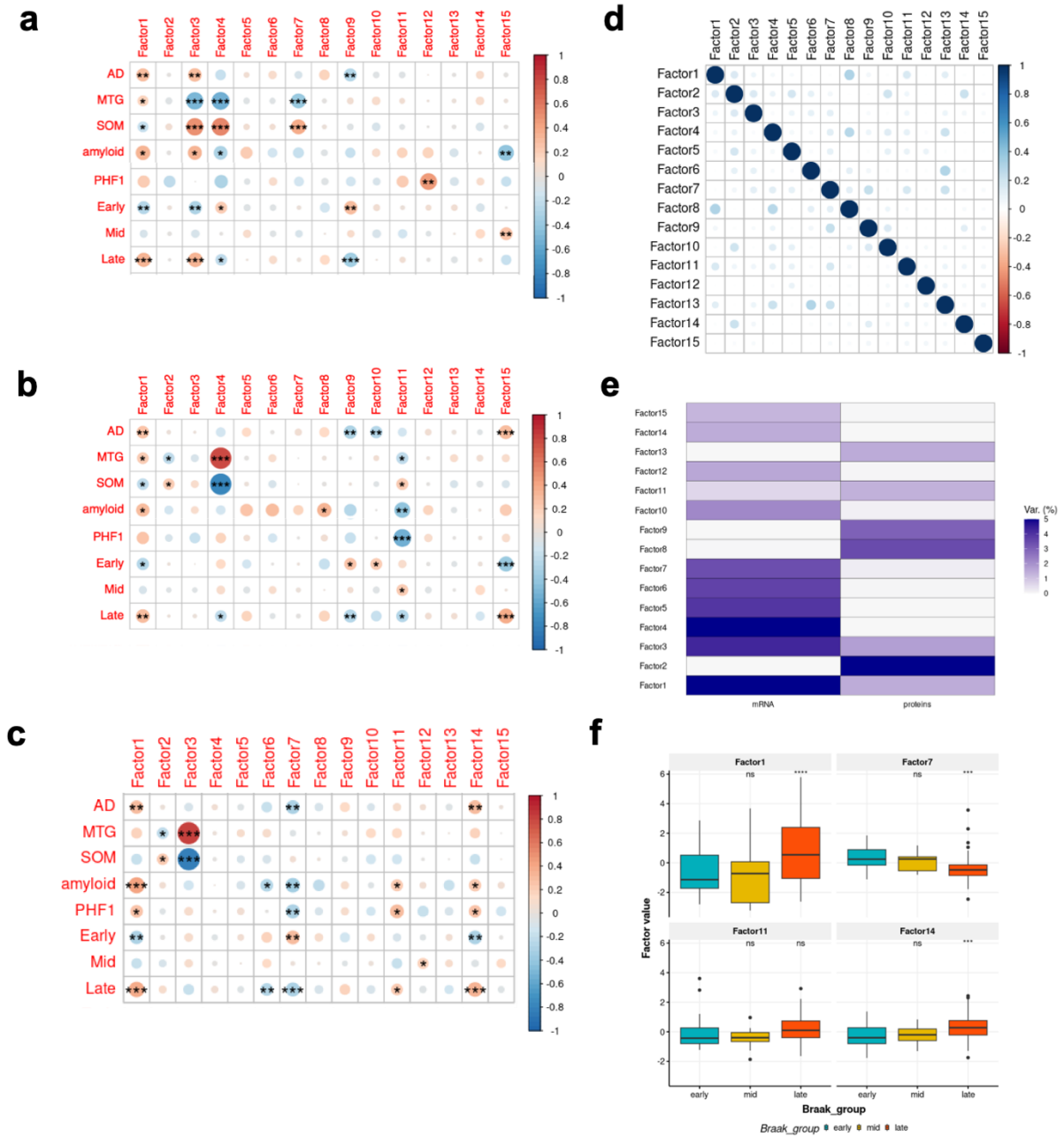


Figure 3-2 Optimisation of MOFA integration

a) MOFA integration option 1. No filtering of features before integration led to unbalanced modalities: 20231 transcripts and 2935 proteins. **b)** MOFA integration option 2. Integration on the intersecting 2756 features and downsize of both proteomics and transcriptomics. **c)** MOFA integration chosen option with balanced view and filtered transcriptomics on the top 14% most variable features. Pearson's correlation heatmap of various clinical and pathological parameters with MOFA factors derived from bulk transcriptomics-proteomics integration. **d)** Factors correlation matrix where poor correlation between factors suggests a good model fit. **e)** Variance explained by the learnt factors across transcriptomics and proteomics. **f)** AD-relevant factors differences between Mid (Braak III-IV) and late (Braak V-VI) relative to early (Braak 0-II) (Two-sided Wilcoxon rank-sum test, adjusted p , * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$).

3.3.2 Cross-sectional pseudotime analysis

These AD-correlated factors (1,7,11, 14) and the brain region-relevant factor 3 were then used in a semi-supervised approach to derive a brain region-dependent disease pseudotime trajectory (**Figure 3-3**). Along this trajectory, pseudotime was correlated with both local tissue phospho-tau (PHF1, $R=0.61$, $p<0.001$) or β -amyloid (4G8, $R=0.56$, $p<0.001$) immunostaining area. I also found that clinical disease progression measured on the Clinical Dementia Rating Scale (CDR)¹⁵⁸ was associated with the pseudotime trajectory. In models where pseudotime, amyloid, or PHF1 were regressed against CDR, the pseudotime model provided the best explanation of variance ($R^2_{\text{pseudotime}}=0.21$, $R^2_{\text{PHF1}}=0.14$, $R^2_{\text{amyloid}}=0.018$).

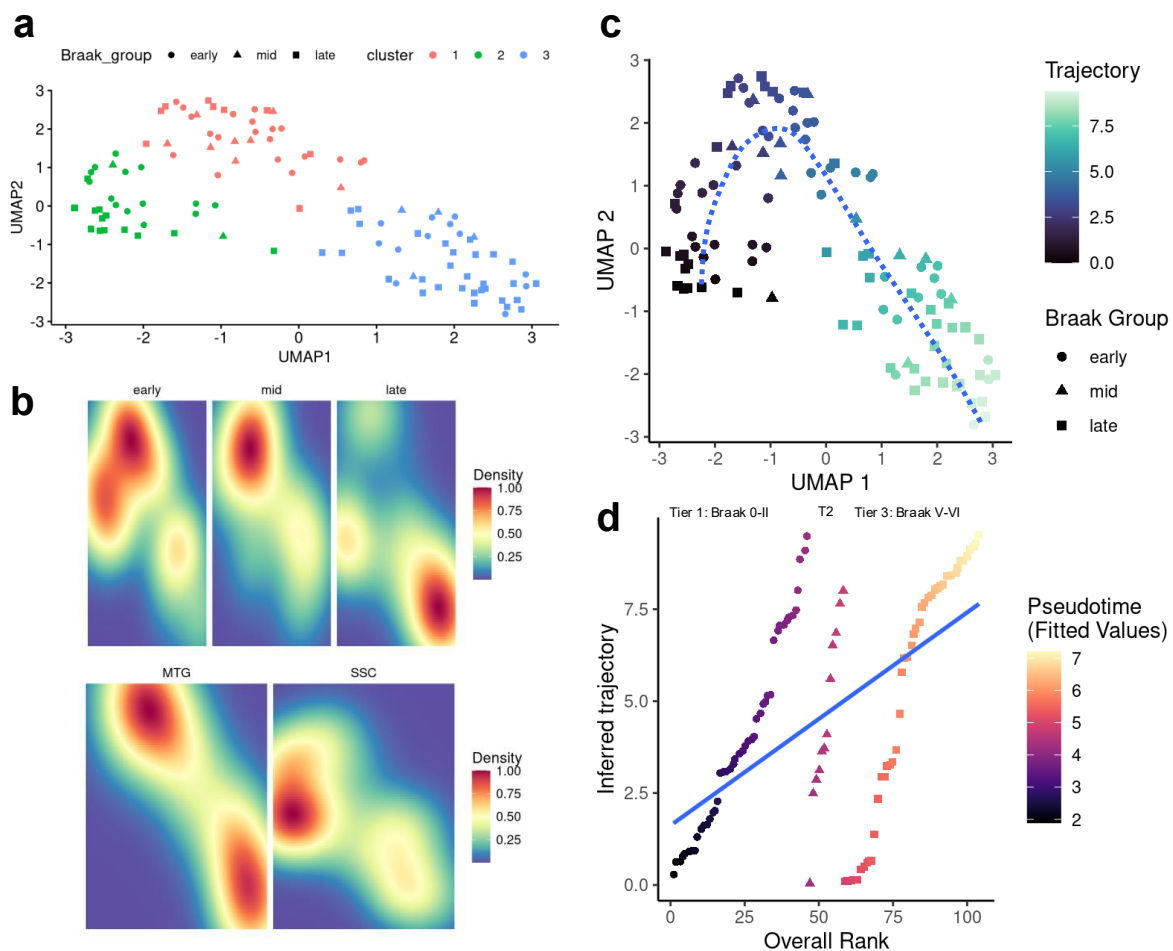


Figure 3-3 Deriving pseudotime from latent factors

a-b) The latent space is clustered into 3 groups via *k*-means clustering. The root of the slingshot trajectory was selected to be the cluster with a high density of early Braak samples (cluster 2) and the ending cluster enriched in later samples (cluster 3) (panel b- top). The root of the

trajectory showed an increased density of SOM samples, which aligns with the SOM being a brain region affected later in AD and, therefore, reflecting early disease changes (panel b-bottom). **c)** Uniform Manifold Approximation and Projection (UMAP) plot illustrating the distribution of brain tissue samples based on a trajectory derived via Slingshot and AD progression. Each point represents a tissue sample coloured by trajectory, ranging from early (black) to late (light cyan) stages of the disease. **d)** This plot illustrates the trajectory of AD progression plotted against the ranked positioning of brain tissue samples according to Braak staging criteria into tiers. The x-axis represents the overall rank of samples from early to late Braak stages, while the y-axis denotes the trajectory inferred through pseudotime analysis. The linear model depicts the correlation between sample ranking within the tiers and disease trajectory. Fitted values correspond to the final pseudotime value.

3.3.3 Transcriptomics-proteomics co-expression networks show AD-correlated glial activation mechanisms and neuronal injury response

I constructed transcriptomics-proteomics co-expression networks based on features contributing to the AD-correlated factors and identified co-expression modules that I associated with AD (**Figure 3-4 a, Supplementary Tables 3-4:5**). I then identified the 15 most highly differentially expressed modules with greater pseudotime (**Figure 3-4 b, Supplementary Tables 3-6:7**). I validated the association of these modules with AD by showing that the eigenvalues of each module were significantly differentially expressed with AD compared to controls in the independent MSBB cohort dataset (**Supplementary figure 3-1 a,b,c,e**). In the ROSMAP cohort, modules didn't replicate, though this was attributed to differences in cohort composition, with most donors with mid-Braak stages III-IV dataset (**Supplementary figure 3-1 a,b,d**).

I used EWCE¹³⁴ to test for their relative enrichment with cell-type data reference from the Allen Human Brain Atlas¹³⁸ to determine brain cell types most strongly associated with each module. Up-regulated modules were enriched for genes most strongly expressed in endothelial cells, oligodendrocytes, astrocytes, and microglia. Down-regulated modules were enriched for genes expressed in GABAergic and glutamatergic neurons (**Figure 3-4 a,b**). I selected three modules with the greatest fold change along pseudotime and the most significant associations with microglia, astrocytes, and neurons: one up-regulated module enriched for microglia gene and protein expression (F14_UP_ME1), one up-regulated module similarly enriched for astrocytes (F14_UP_ME2) and one down-regulated module enriched for neurons (F11_DOWN_ME2). Functional enrichment analysis linked each module to distinct pathology-related pathways. F14_UP_ME1 (microglial) was associated with inflammatory response and synaptic pruning,

F14_UP_ME2 (astrocytic) with metal ion homeostasis, and F11_DOWN_ME2 (neuronal) with synaptic functionality and exocytosis (**Figure 3-4 c, Supplementary files 2-5**). I interpret these as mechanistic pathways expressed with AD with neuropathology and clinical progression (**Supplementary figure 3-2**).

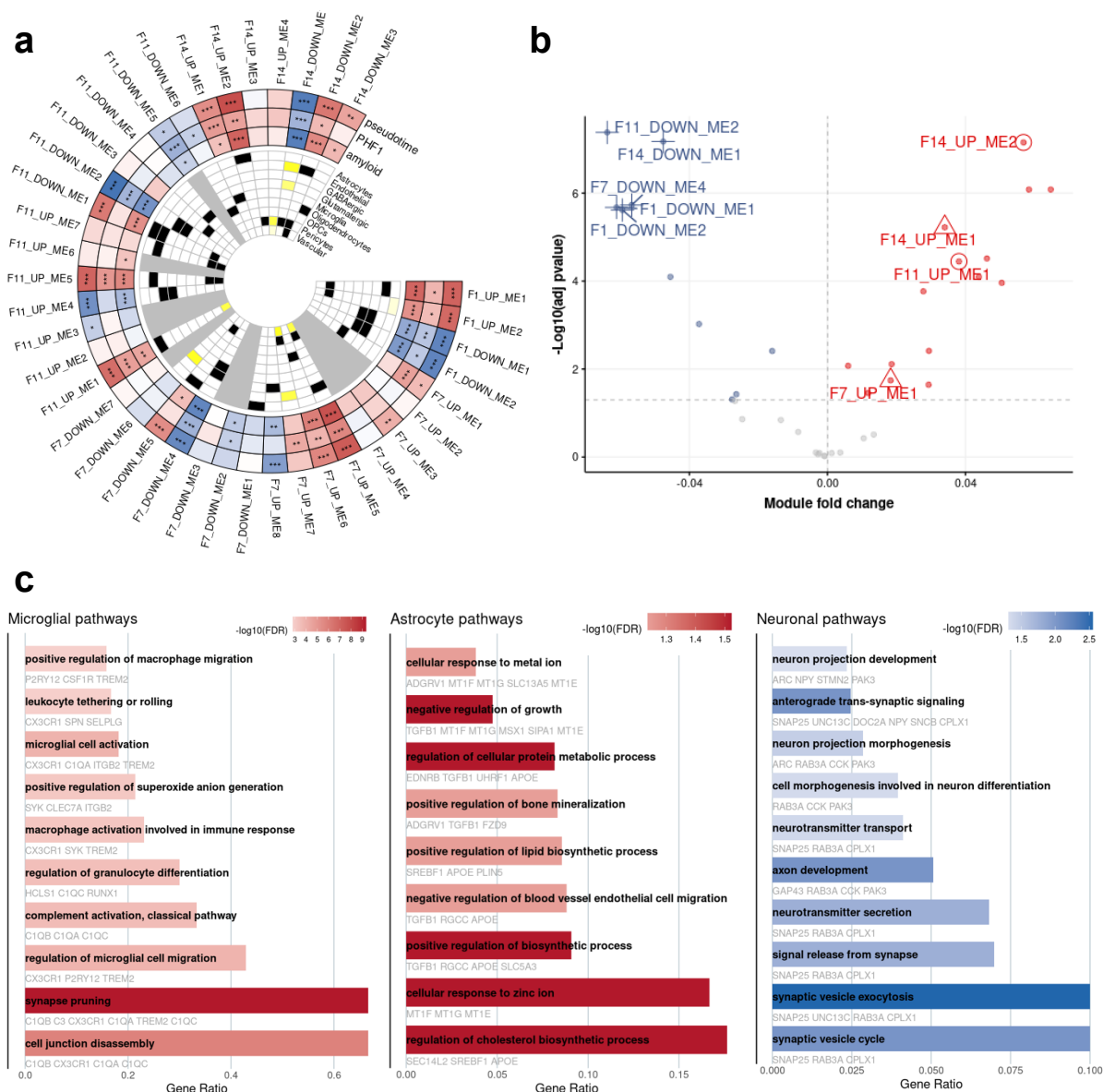


Figure 3-4 Transcriptomics-proteomics co-expression network define AD-correlated glial activation mechanisms and neuronal injury response

a) Circos plot visualising module trait relationships Pearson's correlations indicating significant associations with key AD features (positive association in red, negative in blue). Cell type enrichment in the inner circle indicates significant enrichment in black ($p < 0.05$), near significance in yellow ($0.05 < p < 0.1$), and NS in white. Pearson's correlation p-values derived from t-test ($*p \leq 0.05$, $**p \leq 0.01$, $***p \leq 0.001$). **b)** Volcano plot illustrating module fold changes plotted against pseudotime, accounting for various confounding factors, with up-

regulated (red) and down-regulated (blue) significant modules labelled. Circles, triangles, and crosses indicate astrocyte, microglial and neuronal enrichment, respectively. **c)** Pathway enrichment analysis (GO Biological Process) for selected up-regulated top microglia (F14_UP_ME1) and astrocyte (F14_UP_ME2) modules and the top downregulated neuronal (synaptic) module (F11_DOWN_ME2), with colour intensity representing the $-\log_{10}(\text{FDR})$ and key genes highlighted within each pathway.

3.3.4 Microglia and astrocyte activation mechanisms are cell state-specific and enriched in AD genetic risk contributed either by transcriptomics or proteomics

In previous work from the Matthew's group^{31,185}, snRNA-seq data was previously annotated into distinct cell types whose samples partly overlap with the current bulk cohort (42 out of 112 samples). I focused on microglia, astrocytes and neurons for the present analysis and further sub-clustered cells into cell subtypes. Here, I employed high-dimensional Weighted Gene Co-expression Network Analysis (hdWGCNA) to project the previously identified modules onto the sub-clustered microglia, astrocytes and neuronal snRNA data. This allowed cell and cell sub-type specific expression of genes to be identified within the bulk modules. Specifically, within the microglia population, I identified the Micro2 subtype, which here showed cell subtype-specific expression patterns from the projected microglia bulk modules (**Figure 3-5 a-b**) and DAM and HLA-like states²⁹ (**Figure 3-5 c**). Two distinct astrocyte populations showed cell subtype-specific expression patterns from the projected astrocyte bulk modules, including a “classical reactive response”, and a “mixed reactive response” exhibiting both intermediate reactive and exhausted expression. Indeed, Astro2&5 which corresponded to classical reactive astrocytes (R0/R1/R2) in Serrano-Pozzo clusters, which also showed cell subtype-specific expression patterns from F11_UP_ME1 (**Figure 3-5 e-f**). On the other hand, Astro3&6 which corresponded to intermediate reactive and exhausted astrocytes in Serrano-Pozzo clusters, showed cell subtype-specific expression patterns from F14_UP_ME2 (**Figure 3-5 e-f**). Astro 3&6 (mixed reactive/exhausted) showed higher expression of neurotoxic genes, while Astro2 and Astro5 (classical reactive) exhibited a decreased neurotoxic gene expression signature. Concurrently, Astro2 showed increased neuroprotective gene expression, and Astro5 displayed reduced neuroprotective gene expression, respectively. Therefore, the net effect of the classical reactive astrocyte populations (Astro2&5) is a reduction in neurotoxicity with an overall shift toward neuroprotection, driven by the increased neuroprotective activity of Astro2 and the neutral contribution of Astro5 (**Figure 3-5 g, Supplementary Table 3-8**). In the rest of the analysis, F11_UP_ME1 denotes the classical reactive astrocyte response, while

F14_UP_ME2 denotes a neurotoxic, mixed reactive/exhausted response that may reflect a transitional state between intermediate reactivity and an exhausted astrocyte phenotype. Neurons showed higher expression of downregulated neuronal modules in specific neuronal subtypes, notably including RORB⁺ excitatory subpopulations (*Supplementary figure 3-3*).

Enrichment analyses using MAGMA¹⁷⁴ and AD GWAS^{143,175,176} studies in a meta-analysis were used to assess enrichments for AD genetic risk in microglial and astrocytic bulk modules separately for the transcriptomic or proteomic data alone or for the union of the features. I found that DAM/HLA-associated microglia (F14_UP_ME1/Micro2) signatures were significantly enriched in AD risk genes and that this was most strongly associated with the transcriptomic signatures (*Figure 3-6 a-b*). The proteomics signature for the mixed reactive/exhausted neurotoxic astrocyte module (F14_UP_ME2/Astro3&6) was also enriched for the expression of AD risk genes (*Figure 3-6 a,c*). Through transcription factor (TF) target gene enrichment analysis, I discovered significant enrichment of STAT1 (OR=3.7, p-value=2e-03) and CEBPB (OR= 4.1, p-value= 2e-04) target genes in the microglia module, underscoring their potential as master regulators of the microglial transcriptional response. A similar analysis of the astrocyte module did not identify significant enrichment for transcription factors (*Figure 3-6 d*).

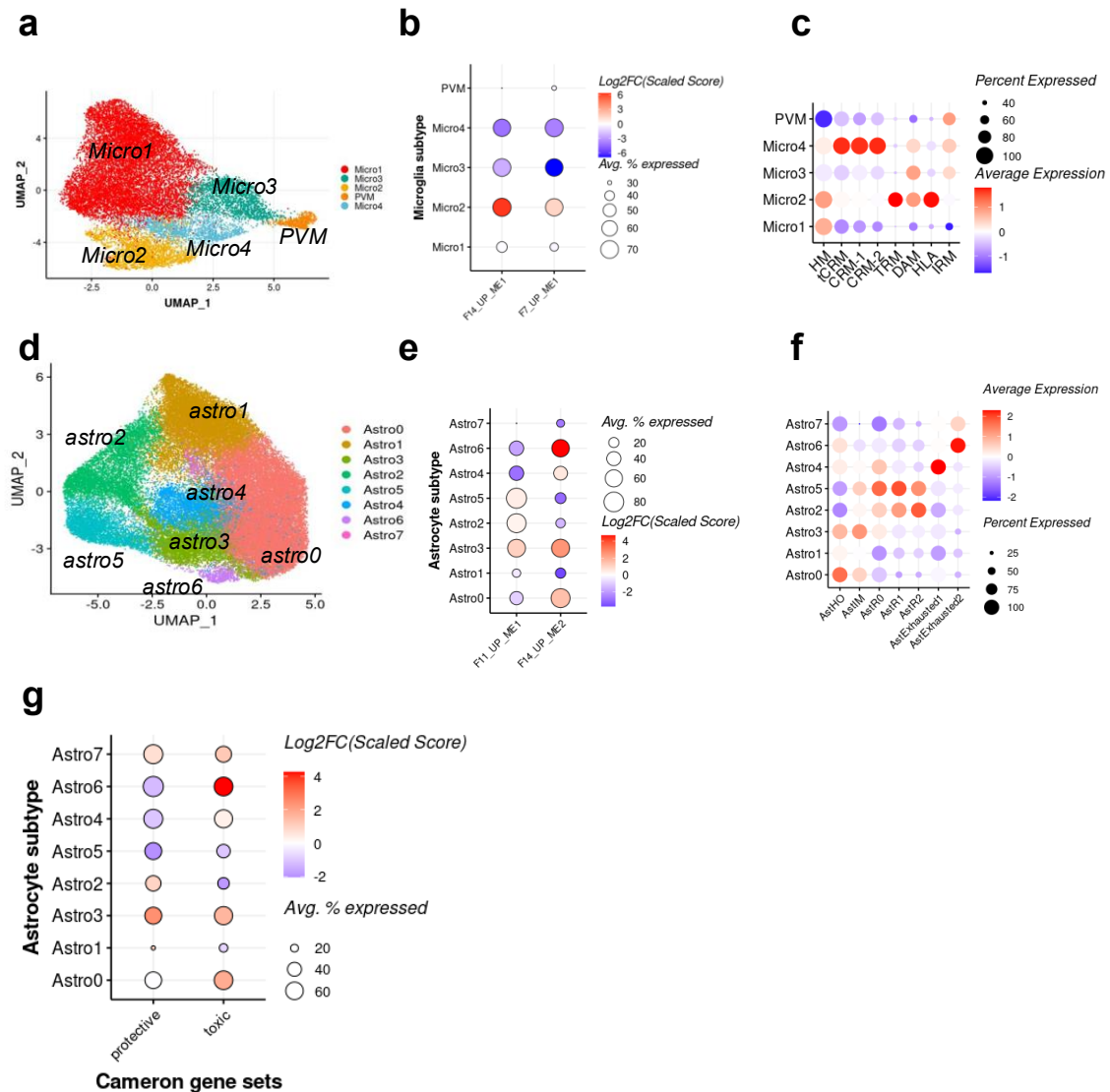


Figure 3-5 Functional characterisation of microglia and astrocyte bulk modules within their sub-cellular context

a) UMAP visualisation of microglial subpopulations in snRNA-seq. **b)** Projecting AD-correlated microglia modules on single-cell clusters using *hdWGCNA* with colour intensity indicating the Log_2FC in module score between AD patients and NDC. The *micro2* cluster is specific to the two microglial-enriched bulk modules, *F14_UP_ME1* and *F7_UP_ME1*. Correspondence is assigned to the module-subpopulation pair with the most significant average expression and greatest difference in module score in AD vs. Controls. **c)** Dot plots quantifying the average percentage expression of Mancuso et al. defined microglial subtypes²⁹. Colour gradient representing scaled average expression levels. **d)** UMAP visualisation of astrocyte subpopulations in snRNA-seq. **e)** Same as in panel b) but using astrocyte-correlated modules on astrocyte subpopulations. *Astro2&5* (classical reactive) and *Astro3&6* (Mixed reactive/exhausted) correspond to *F14_UP_ME2* and *F11_UP_ME1* respectively. **f)** Dot plots quantifying the average percentage expression of Serrano-Pozo et al. defined astrocyte subtypes²⁹. Colour gradient represents the scaled average expression levels. **g)** Dot plots quantifying the Log_2FC in module neuroprotective/neurotoxic score between AD patients and NDC by clusters. Genesets derived from Cameron et al. astrocyte neuroprotective and neurotoxic subtypes¹⁶².

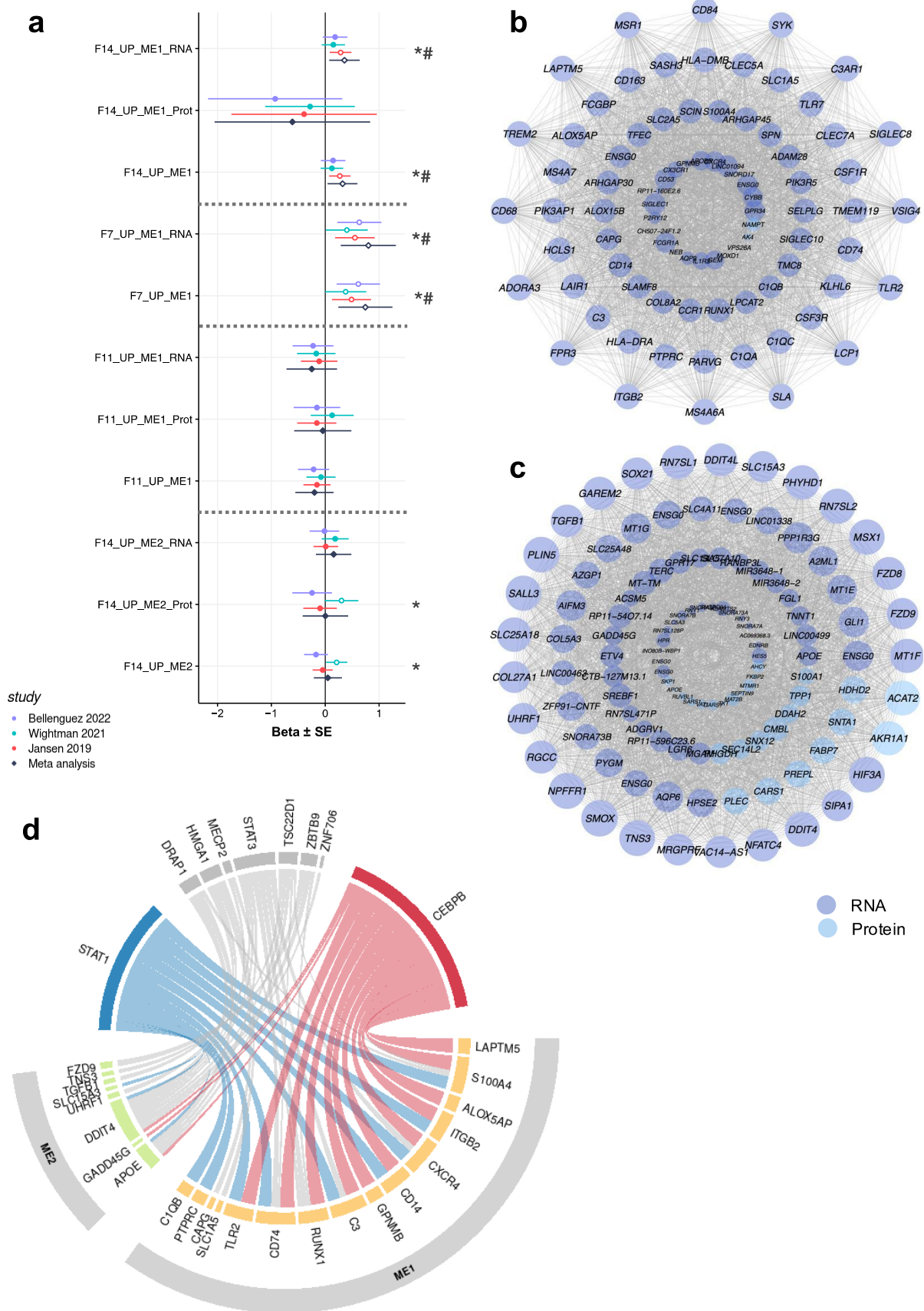


Figure 3-6 Microglia and astrocyte activation mechanisms are enriched in AD genetic risk
a) Forest plot illustrating the enrichment of microglia (F14_UP_ME1, F7_UP_ME1) and astrocyte co-expression modules (F14_UP_ME2, F11_UP_ME1) (split by transcriptomics ‘_RNA’, proteomics ‘_Prot’, and union of features) in genetic risk quantified by three AD GWAS^{143,175,176} and in a meta-analysis, highlighting significant beta values and corresponding P-values. # indicates modules enriched in AD risk genes in the meta-analysis and * only in one or more studies. **b-c)** Schematic co-expression network diagrams showing module membership of genes (dark blue) and proteins (light blue) in F14_UP_ME1 (top) and F14_UP_ME2 (bottom). Node size is proportional to connectivity (degree), and nodes are separated into four quantiles representing each layer. **d)** Integrated transcriptomics-proteomics analysis uncovers disease progression-specific TF regulatory mechanisms in microglia and astrocytes. Circos plot displays TF proteins and their target genes in ME1 (F14_UP_ME1) microglia and ME2 (F14_UP_ME2) astrocytic modules. TFs with significant enrichment of target genes in ME1 are shown in blue (TF with negative factor 14 weight) and red (TF with positive factor 14 weight). Other TF with no significant enrichment of target genes in modules are indicated in grey.

3.3.5 AD polygenic risk modulates glial activation and its effects on the rate of cognitive decline

To explore the genetic influence on glial activation and clinical disease in an early AD cohort, I computed polygenic risk scores (PRS) derived from SNPs in genes within microglial and astrocyte activation modules that I had shown to be enriched for AD risk (**Figure 3-7, Supplementary Tables 3-9:11**). PRS scores were computed for the subset of data from donors in the ROSMAP cohort with early AD (Braak stages III-IV, $n=132$). Although pathway-specific methods such as PRSet could offer complementary insights, I opted for a locus-based approach to directly integrate with disease modules. While no formal power calculations were performed, I aimed to improve signal detection by restricting SNPs to those mapped to genes driving the genetic enrichment.

To determine the optimal PRS for predicting phenotypic outcomes, I conducted regression analyses across various p-value thresholds (from 10^{-7} to 0.5) against two phenotypic variables: global AD pathology and microglia/astrocyte modules eigenvalues. The best-fit PRS was selected based on the model that explained the highest proportion of phenotypic variance. The optimal model fit for the Microglia PRS was achieved at a p-value threshold of 0.001, which included 68 SNPs (**Figure 3-7 a-c**). This model maximised the R^2 value when modelled against microglia module eigenvalue, indicating predictive capability. Of these, 61 SNPs were present in the ROSMAP genomic data. The best fit for the Astrocyte PRS was found at a p-value

threshold of 0.005, encompassing 71 SNPs, with 21 detectable within the ROSMAP genomic data (**Figure 3-7 b-d**). This model demonstrated the highest R^2 when correlated with global pathology measures. These findings are detailed further in **Supplementary Tables 3-10 and 3-11**, which include comprehensive lists of the SNPs included in each PRS scores after thresholding.

The microglial PRS was significantly correlated with microglial module eigenvalue, consistent with genetic influence on microglial activation (**Figure 3-7 e**). Despite this genetic influence on microglial module eigenvalue, microglial PRS did not significantly correlate with neuropathological measures or the rate of clinical progression. Conversely, the astrocyte PRS, featuring SNPs from genes including *PLEC* and *APOE* (**Figure 3-7 b-d**), did not directly correlate with astrocyte module eigenvalue but was correlated with A β burden (expressed as amyloid load, diffuse or neuritic plaques counts) (**Figure 3-7 e**). An increased neurotoxic mixed reactive/exhausted response (F14_UP_ME2 eigenvalue) was significantly associated with neuropathology and a faster rate of cognitive decline, whereas the classical reactive astrocyte response (F11_UP_ME1 eigenvalue) showed no such association in this early AD cohort (**Figure 3-7 e**). Moreover, neurotoxic mixed reactive/exhausted astrocyte and microglia eigenvalues were positively associated with pseudotime (**Figure 3-7 e**). Together, these results suggest that genetic risk is a determinant of microglial activation in AD. By contrast, my findings indicate that astrocyte reactivity in AD involves two distinct responses. The classical reactive response shows no genetic association and occurs before widespread pathology develops, representing an initial response. In contrast, the neurotoxic mixed reactive/exhausted response, which shows a low genetic association, is tightly linked to advanced neuropathology and cognitive decline, suggesting it is primarily driven by local signals such as amyloid burden and neuroinflammation.

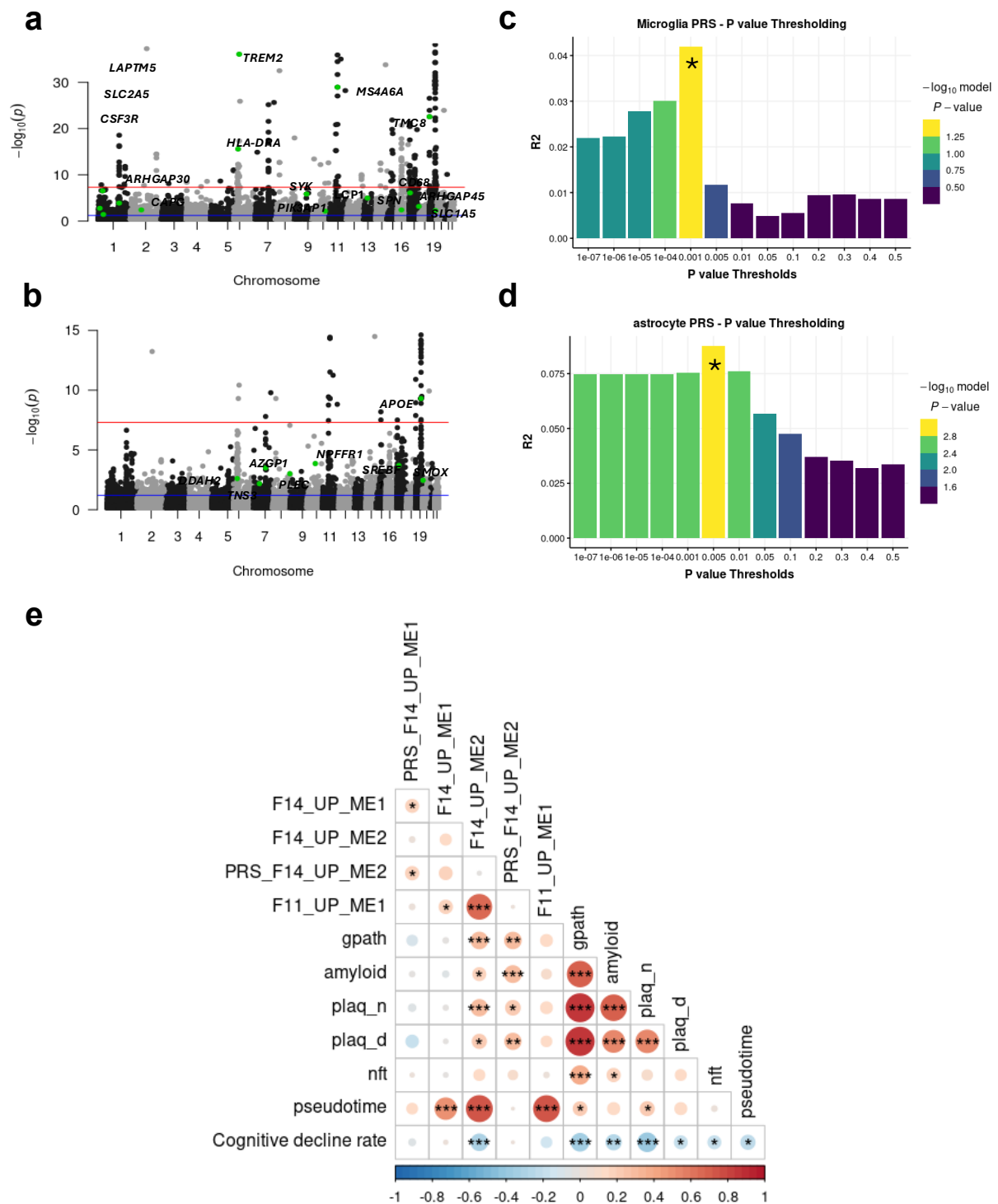


Figure 3-7 AD polygenic risk modulates glial activation directly and indirectly

a-b) Manhattan plots displaying the MAGMA gene-based association test results for AD-correlated modules F14_UP_ME1 (microglia) and F14_UP_ME2 (astrocyte), with genome-wide significance thresholds indicated and notable genes highlighted. The red line indicates genome-wide significance ($p \leq 5 \times 10^{-8}$) and the blue line ($p \leq 0.05$). **c-d)** P-value thresholding according to global pathology. Astrocyte PRS R^2 was maximised when modelled against global pathology, with a best-fit PRS at the 0.005 threshold (denoted by *). P-value thresholding according to microglia/astrocyte module eigenvalues. Microglia PRS R^2 was maximised when

modelled against microglia module activity with a best-fit PRS at the 0.001 threshold (denoted by *). **e)** Pearson's correlation heatmap of microglia and astrocyte PRS, modules, and neuropathology measures, including a global pathology measure (quantitative summary of AD pathology derived from counts of three AD pathologies: neuritic plaques (n), diffuse plaques (d), and neurofibrillary tangles (nft), and rate of cognitive decline in the last three years preceding death. Pearson's correlation p-values derived from t-test (* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$).

3.3.6 Glial activation is linked to synapse dysfunction and a greater rate of cognitive decline in early AD

I clustered individuals from the early AD ROSMAP cohort (Braak III-IV stages, $n=132$) via unsupervised multi-omics clustering of transcriptomics and proteomics¹²⁷ to identify potential molecular subgroups in early AD. This revealed two molecular subgroups that showed no difference in demographics or neuropathology (**Figure 3-8 a-b, Supplementary Tables 3-12:13**). In parallel, I assessed survivor bias, indicating that patients with follow-up years >12 should be excluded from the following Kaplan-Meier survival analyses (**Supplementary figure 3-4**).

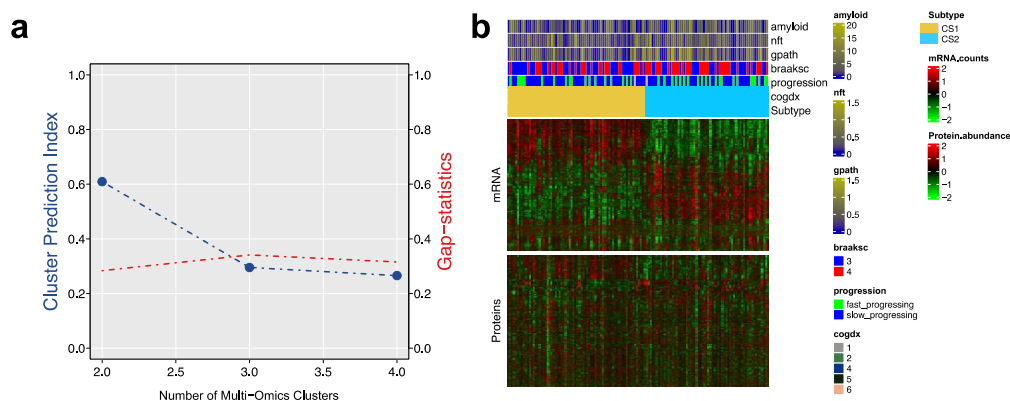


Figure 3-8 Early AD ROSMAP cohort stratification via multi-omics clustering

a) Cluster number optimisation: The optimal number of clusters was determined by combining the CPI and Gap-statistics results, selecting the number that maximised the sum of the CPI score and Gap statistic ($k=2$). **b)** Multi-omics clusters transcriptomics and proteomics expression heatmap.

I tested whether the glial activation mechanisms identified in astrocytes and microglia — enriched for AD risk genes — and the neuronal injury response module, were associated with accelerated cognitive loss in this early-stage AD cohort (i.e. which comprises two molecular subgroups as described above). Kaplan Meier analysis revealed that the two molecular subgroups identified via unsupervised multi-omics clustering mapped with faster (blue) and

slower (yellow) clinical progression (**Figure 3-9 a, Supplementary Table 3-13**). I also analysed the global cognitive function, which is derived from the average of 19 cognitive tests. These tests span multiple domains, including episodic, semantic and working memory, visuospatial ability and perceptual speed. Most ROSMAP participants have taken these tests annually, allowing for a detailed longitudinal assessment of cognitive changes. When examining longitudinal cognitive profiles, I observed significant differences in the rate of cognitive decline between these groups, where the fast group showed an accelerated rate of decline compared to the slow group (**Figure 3-9 b**). Further downstream analysis uncovered that it was primarily driven by the accelerated rate of decline in episodic and working memory and not by perceptual speed, visuospatial ability, and semantic memory (**Supplementary Table 3-14**). Conducting a sensitivity analysis with an alternative assessment of asymptomatic vs. symptomatic AD using the MMSE⁶⁰ value indicated that the fast cluster exhibited odds of 2.85 times higher for symptomatic presentation than the slow group (**Supplementary figure 3-4**). Furthermore, modelling using a random forest model trained on AD bulk modules to predict pseudotime in the primary cohort predicted higher pseudotime values in the fast group compared to the slow group (**Figure 3-9 c**). Feature importance analysis revealed that this prediction was predominantly driven by the neuronal injury response (F11_DOWN_ME2) and reactive astrocyte modules (F14_UP_ME2) (**Figure 3-9 d, Supplementary figure 3-5**). I found statistically significant differences in module eigenvalues between the two groups, with the fast group exhibiting higher glial activity and more pronounced neuronal injury response (**Figure 3-9 e**). Projecting the genes and proteins up regulated in the fast versus the slow group (**Supplementary Table 3-15**) onto the discovery cohort snRNA dataset revealed distinct expression patterns. Genes and proteins up regulated in the fast progression group showed high expression in the Micro2 and Astro3&6 clusters, the latter corresponding to the neurotoxic mixed reactive/exhausted astrocyte states. In contrast, genes and proteins up regulated in the slow progression group were enriched in the Astro2 cluster (**Figure 3-9 f**), representing the classical reactive response, which have previously been associated with a neuroprotective signature in **Figure 3-5**.

To test for synaptic differences between the groups, I leveraged a SNARE proteomics dataset acquired on the same ROSMAP participants, quantifying synaptic protein-protein interaction of presynaptic components involved in the SNARE complex (i.e. known to mediate exocytosis)

via heterologous capture ELISA¹⁸⁰. I found that the level of SNAP25-VAMP interaction was diminished in the faster-progressing group ($p < 0.05$). This reduction was not paralleled in other synaptic components, highlighting a specific synaptic dysfunction in SNAP25-VAMP interaction in the fast group (**Supplementary Table 3-16**). Therefore, one can reverse the scale whereby higher SNAP25-VAMP dysfunction corresponds to lower levels of interaction. A linear mixed-effects model revealed a significant interaction between SNAP25-VAMP dysfunction and cluster group classification (slow vs. fast), demonstrating that the relationship between SNAP25-VAMP functionality and the rate of cognitive decline varies across groups. Specifically, the fast-progressing group exhibited a more pronounced decline associated with SNAP25-VAMP dysfunction than the slow-progressing group, highlighting group-specific vulnerabilities to synaptic impairments (**Figure 3-9 g, Supplementary Table 3-19**). I found that the expression of the neuronal injury response module, F11_DOWN_ME2, was strongly associated with SNAP25-VAMP interactions, highlighting that the module captures variation related to synaptic dysfunction in AD ($p = 0.004$, **Supplementary Table 3-17**). SNAP25-VAMP dysfunction was positively associated with the neurotoxic mixed reactive/exhausted astrocyte module in the fast group but not with the microglia module, nor the classic astrocyte reactivity module (**Figure 3-9 h, Supplementary tables 3-19:20**). These findings highlight an association between a neurotoxic astrocyte response, synaptic dysfunction (i.e. synaptopathy), and rate of cognitive decline.

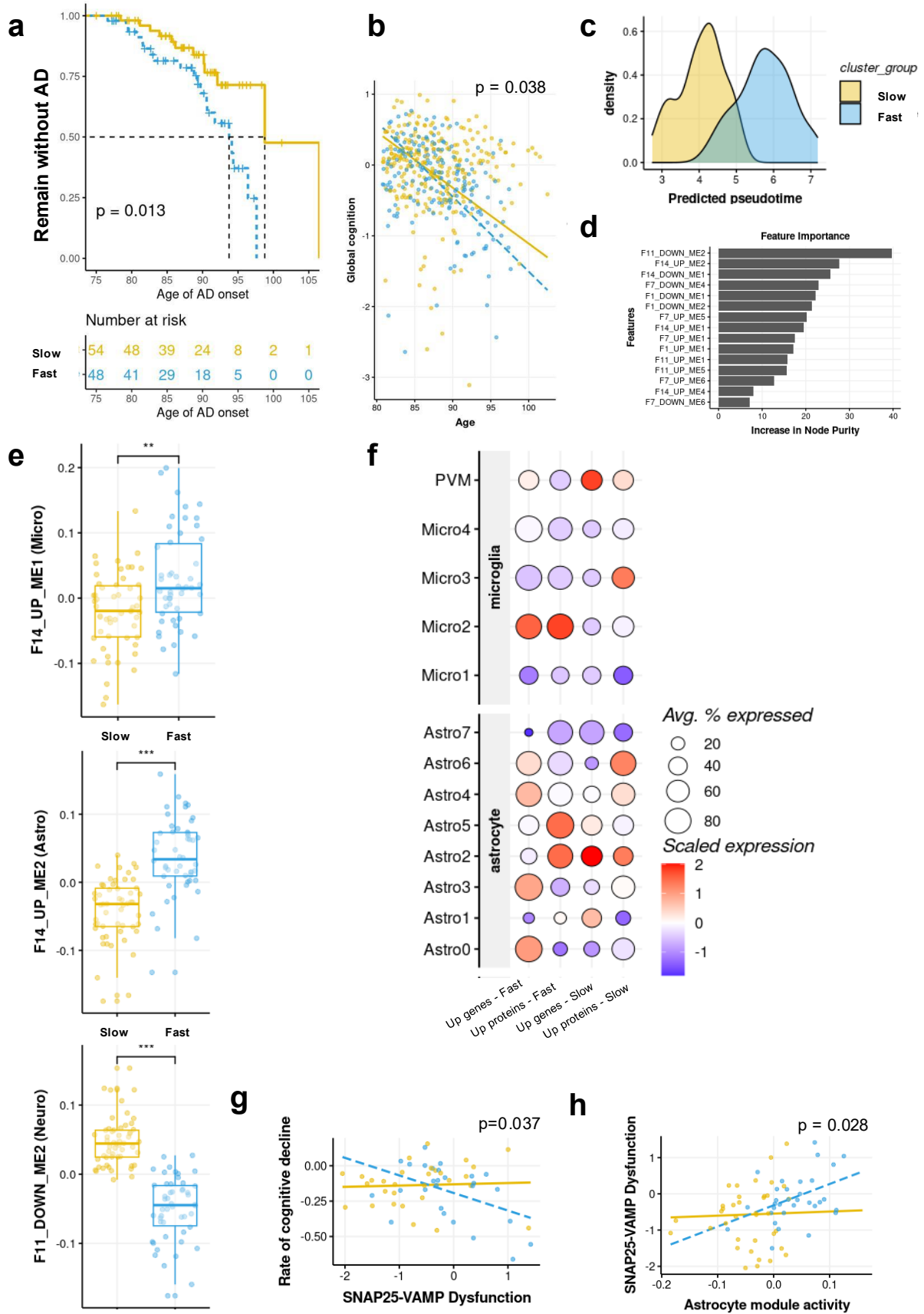


Figure 3-9 Identified glial activation mechanisms are related to the trajectory of cognitive decline and synapse dysfunction

a) Kaplan-Meier curves represent a delay in AD onset. Survival is quantified using longitudinal cognitive assessments, in which non and mild-cognitively impaired individuals without an AD diagnosis are included (preclinical and prodromal progression to dementia). The analysis is stratified by subgroups identified in multi-omics clustering analysis and identifies fast (blue) and slow (yellow) groups based on their differential cognitive loss profiles. **b)** Linear mixed model assessing loss of global cognition as a function of age, highlighting a significant differential decline over time. **c)** Density plot depicting the predicted pseudotime in early AD, comparing slow and fast groups. The random forest model was trained on the discovery cohort AD-correlated bulk modules, and the trained model was applied to the validation cohort to predict pseudotime. **d)** A bar chart showing the importance of bulk modules, ranked by increasing node purity, in predicting pseudotime. **e)** Box plots comparing module scores between slow and fast patients. (Wilcoxon rank-sum test, adjusted p , $*p \leq 0.05$, $**p \leq 0.01$, $***p \leq 0.001$). **f)** Dot plots showing the expression of up and down-regulated genes and proteins between fast versus slow groups (Supplementary Table 3-15) across microglial subtypes and astrocyte subtypes, with size representing the average percentage expressed and colour intensity denoting the difference in module score between Fast versus Slow groups. **g)** SNAP-VAMP25 dysfunction is associated with accelerated cognitive decline in the fast group. (i.e. negative rate indicates a decline in cognition). **h)** Synaptic dysfunction is associated with neurotoxic astrocyte module activity only in the fast group.

3.3.7 Biomarkers in plasma and CSF: Identifying high-risk groups for rapid cognitive decline

To test the translational relevance of the co-expression modules, I investigated their feature component in CSF and plasma to identify potential biomarkers for identifying individuals at high risk for rapid cognitive decline in AD. I used plasma and CSF proteomics modules identified in a previous study from Dammer et al.¹⁸² and tested overlap with my candidate modules using Fisher's exact tests. Module nomenclature and annotations are taken from the original study¹⁸²; details can be found in the methods section. For each pair of modules—one from Dammer et al. and one from my analysis, I tested whether the number of shared genes or proteins was greater than expected by chance. Modules were considered overlapping if the enrichment was statistically significant after multiple testing correction. In CSF, biomarkers were identified for microglial activation and neuronal injury response modules but not for the reactive astrocyte module (**Figure 3-10**). Notably, the overlap between Dammer's CSF microglial modules M7 and F14_UP_ME1 includes proteins such as C3, CLEC7A, CD163, VPS26A, and VSIG4. For synaptic dysfunction, there was a significant enrichment of F11_DOWN_ME2 in both M6 and M8, with GAP43, STMN2, and NEFL being the overlapping proteins. These synaptic biomarkers correlated positively with CSF pTau and

showed up-regulation in AD CSF versus control. In plasma, biomarkers for microglial activation predominantly involve the complement system, with modules M28 and F14_UP_ME1 showing significant enrichment for the complement proteins C3, C1QA, C1QB, and C1QC. SNAP25 was detected in plasma as part of the M2 module, although the enrichment test with the neuronal injury module did not reach significance.

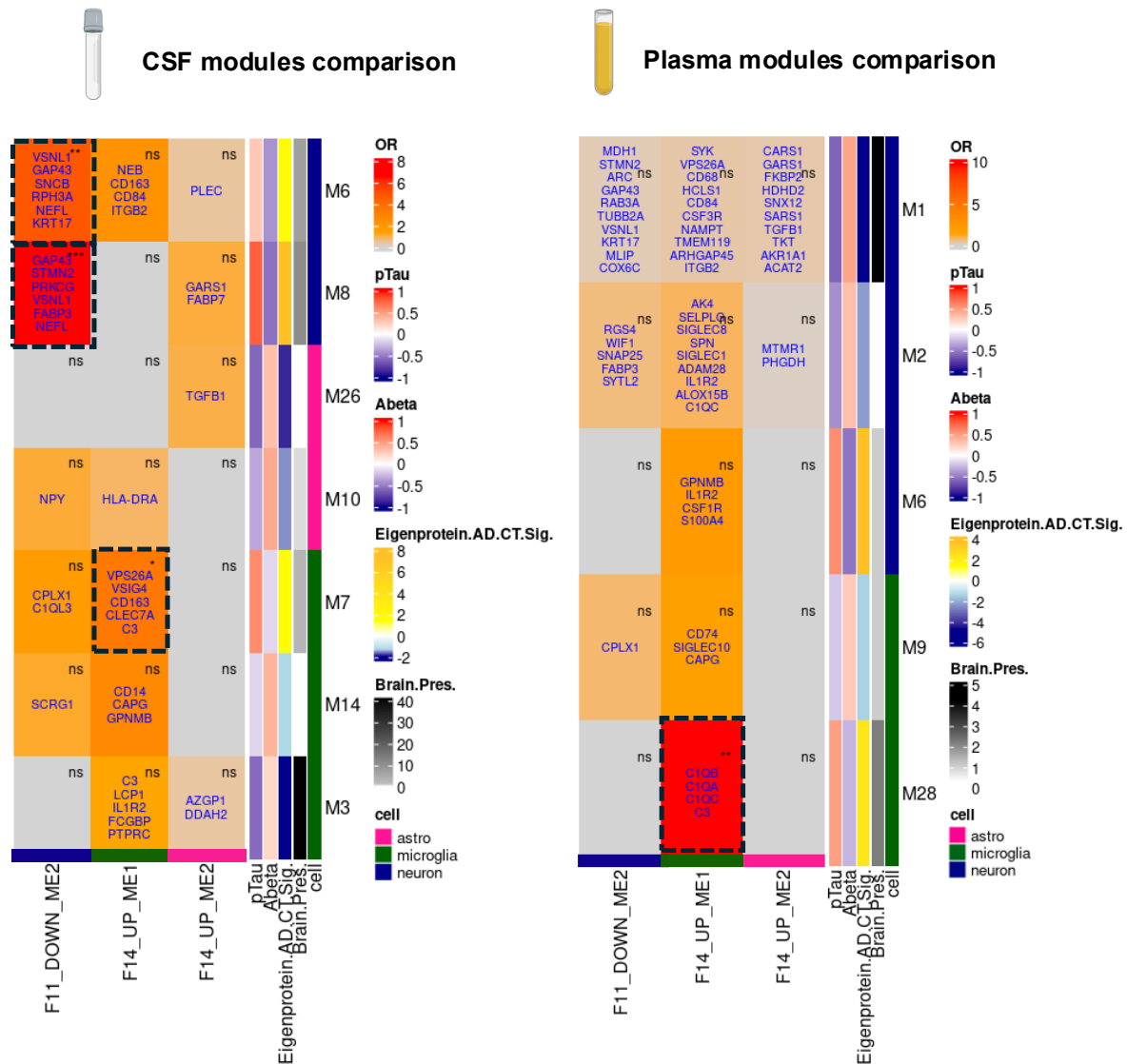


Figure 3-10 Biomarkers of glial activation and synapse dysfunction in CSF and plasma as potential biomarkers for Alzheimer's accelerated progression

Reference CSF and plasma proteomics co-expression modules were taken from Dammer et al.¹⁸². Modules that were significantly differentially expressed in AD versus control and enriched for neurons, astrocytes or microglia were kept for analysis. OR: Fisher's Exact test Odds Ratio. (Fisher's exact test, * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$). Gene names in blue overlap with the Dammer et al. modules compared to this study. The black dotted boxes show module pairs significantly overlapping between internal and Dammer et al. modules. All the row

annotations were taken from Dammer et al. supplementary material¹⁸². Briefly: pTau and Abeta are derived from CSF and show Pearson's correlation values with the module eigenvalue. Brain Pres shows the module's brain preservation $-\log_{10}(p\text{-value})$, which measures how well a protein module identified in the brain is conserved in CSF or plasma. Eigenprotein AD.CT.sig shows signed $-\log_{10}(p\text{-value})$ of module eigenvalue in AD vs. Controls. Cell colour coding pink, microglia and blue shows significant module enrichment for astrocytes, microglia, or neurons in the original publication.

3.3.8 Linking the rate of cognitive decline to the notion of cognitive resilience

I aimed to investigate whether the selected candidate biomarkers have predictive value in stratifying individuals by cognitive resilience status. For this purpose, I used a CSF proteomics dataset from an independent cohort¹⁸⁴. Cognitive resilience is characterised by individuals with AD pathology exhibiting less cognitive impairment than expected¹⁸³. In this cohort, I categorised individuals as either cognitively resilient or vulnerable according to expected pathology levels based on a regression analysis of cognitive scores against pathology measurements. Cognitive scores were measured with the Montreal Cognitive Assessment (MoCA), where higher scores represent better cognitive function. In **Figure 3-11** (a-c), I regressed the MoCA cognitive score against three CSF pathological measures, tTau, pTau, and A β 42, to assign a continuous cognitive resilience score as previously proposed¹⁸³. This score is defined as the difference between observed cognition and the cognition predicted by the CSF levels of these pathology markers. As expected, pTau and tTau were negatively associated with MoCA scores, while A β 42 exhibited a positive association, as expected from the literature¹⁸⁶. I assigned a resilient status to individuals who had positive cognitive resilience scores for all three pathology measures, resulting in four groups: Resilient AD (n=42), Vulnerable AD (n=107), Resilient Control (n=104), and Vulnerable Control (n=43).

Figure 3-11 (d-f) explores the relationship between the cognitive resilience scores (based on tTau, pTau, and A β 42) and the neuronal, microglia, and astrocyte biomarker scores, stratified by AD status. To capture the balance between neuroinflammatory glial activation and neuronal integrity, I computed a neuro-glia score as the difference between neuronal and glial biomarker scores. I found significant associations between all biomarker scores and pTau/tTau cognitive resilience in individuals with AD (near significance with A β 42 CR), but not in controls. Specifically, I observed a positive relationship between the neuronal score and cognitive resilience, indicating that lower levels of GAP43/NPY were associated with decreased

resilience. Conversely, there was a negative association between the glial biomarkers (C3, CD14, and AZGP1) and resilience, indicating that higher levels of these markers corresponded to decreased cognitive resilience. Reduced glial contribution, reflected by a higher neuron-glia difference, was associated with decreased cognitive resilience in controls, but only when resilience was computed using A β 42 levels and not tau **Figure 3-12** (d-f). As A β 42 accumulation is an early event in AD, often preceding significant cognitive decline or neuronal damage, this finding likely reflects a protective glial response to early amyloid burden in controls and thus suggests that glial activation may play a compensatory role in maintaining cognitive resilience during the initial stages of pathology.

As a sensitivity analysis, disease modules in the early AD ROSMAP cohort studied in section 3.3.6 were tested for association with cognitive resilience and vulnerability using the same approach, which further supported the role of F14_UP_ME2 module (neurotoxic mixed reactive/exhausted astrocyte) and F11_DOWN_ME2 (synaptic dysfunction) in cognitive vulnerability (**Supplementary figure 3-6 a,b,c**). Unexpectedly, F14_UP_ME1 (microglia) showed no association with cognitive resilience, likely due to stage-specific effects (**Supplementary figure 3-6 d**). In other words, the fast/slow classification from section 3.3.6 can be partially linked to the cognitive resilience classification.

In **Figure 3-12** (a-b), I present the ROC curves of the random forest models used to predict cognitive resilience status from tTau, pTau, A β 42, age and sex without and including the neuro-glia biomarker score. I calculated the AUC for each class using a one vs. rest approach. The addition of the neuro-glia biomarkers improved classification performance, particularly for AD resilience and vulnerability, with the AUC increasing from 0.76 to 0.83 for AD-resilient (DeLong's test p-value= 0.002), and from 0.84 to 0.88 for AD-vulnerable (DeLong's test p-value= 0.01). Predictions for controls remained consistent between the two models.

Next, in **Figure 3-12** (c-d), I applied both trained models to an independent dataset of 96 individuals, comprising 33 AD patients, 31 asymptomatic AD individuals, and 32 controls. I kept the original status assignment where asymptomatic AD was defined by CSF levels of amyloid- β , total tau and phospho-tau consistent with an AD diagnosis, but without cognitive impairment. For model performance assessment, I grouped together all predicted controls and

AD individuals regardless of resilient/vulnerable to obtain two groups. Similarly, for the observed groups, I grouped asymptomatic AD and AD into a single AD class. The model, including neuro-glia biomarkers, marginally improved accuracy (0.83 vs. 0.82) and sensitivity (0.75 vs 0.73).

To further dissect these models' predictions, I compared individual-level label assignments. Most labels (74/96) agreed between the two models (*Figure 3-12 e*). Among the observed AD group, 50% (17/33) were consistently predicted as AD_vulnerable by both models. In the observed controls, 48% (16/32) were predicted as Control_resilient across both models. Asymptomatic AD individuals were predominantly split into Control_resilient (13/31) or AD_vulnerable (11/31).

For the remaining 22% of labels, Model 2, which included the additional biomarker score, predicted 12 individuals to have “worse” statuses (vulnerable instead of resilient) and 10 individuals to have “better” statuses (resilient instead of vulnerable) compared to Model 1 (*Figure 3-12 e*). This raises the possibility that the neuro-glia model is capturing additional biological information that may better differentiate patients. Tracking the cognitive trajectories of these individuals prospectively (i.e. when the data becomes available) will be critical to validate whether those predicted to be resilient indeed experience slower cognitive decline compared to those predicted to be vulnerable. This will help determine whether the 22% difference in label assignment with Model 2 aligns more closely with actual disease progression and that model 2 with the neuro-glia score can better predict future trajectories.

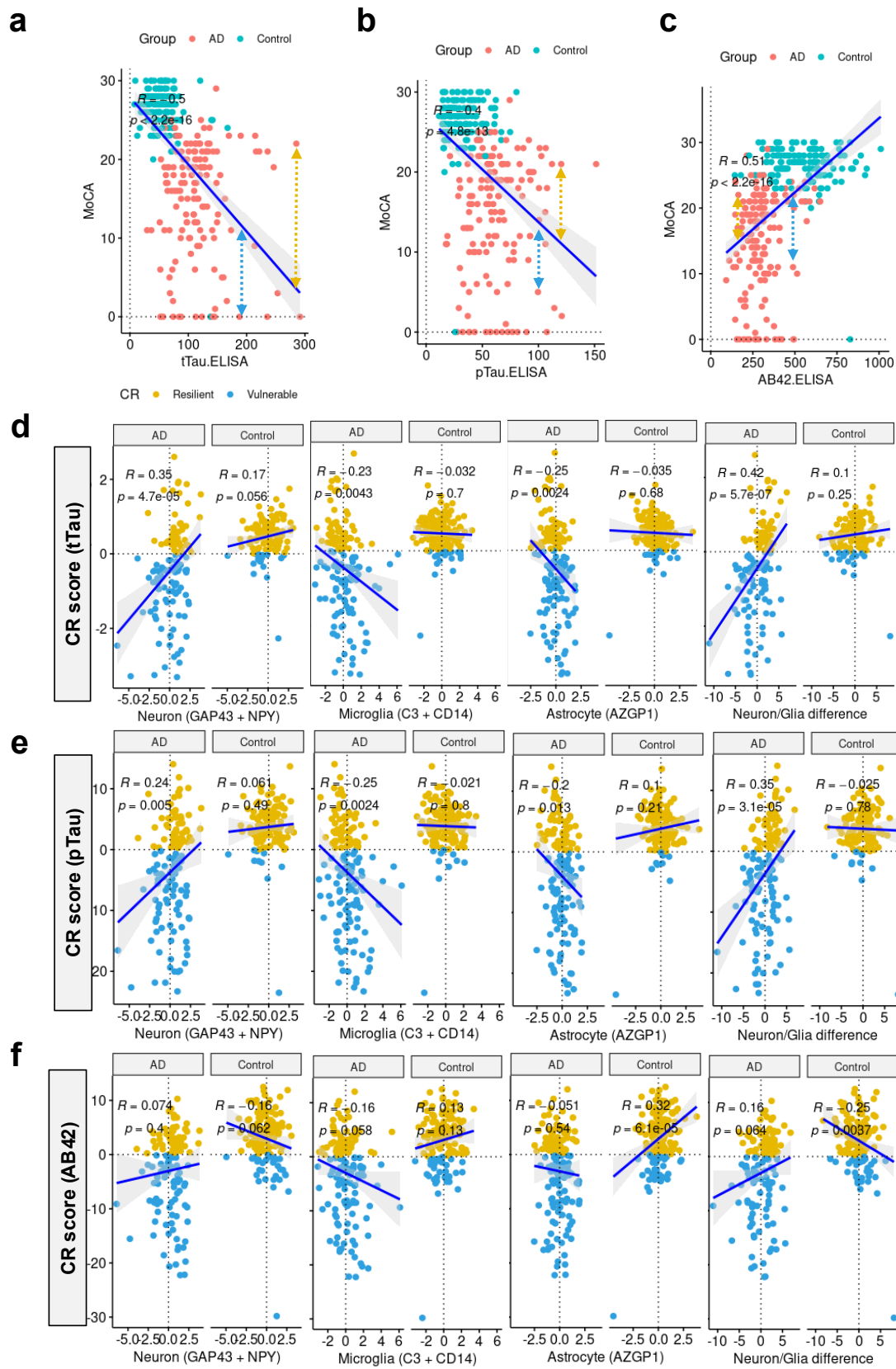


Figure 3-11 Defining the cognitive resilience score

a-c) Linear regression plots showing the relationship between MoCA cognitive scores and CSF pathological markers (tTau, pTau, and A β 42) in individuals with AD (red) and controls (blue). Regression lines demonstrate the negative association of tTau (a) and pTau (b) with MoCA scores, and a positive association of A β 42 (c) with MoCA. Individuals were classified as cognitively resilient (CR > 0 across all pathologies, yellow arrows) or vulnerable (blue arrows). The MoCA values are shown as a continuous variable across the pathology levels. d-f) Scatter plots showing the association of cognitive resilience (CR) scores with various biomarker scores across AD and control groups. Panels (d) represent the CR score based on tTau, (e) for pTau, and (f) for A β 42. Each column shows the relationship between CR and neuronal (GAP43 + NPY), microglia (C3 + CD14), astrocyte (AZGP1) biomarker scores, and the neuro-glia difference (Neuron - Glia). Positive correlations between neuronal markers and CR scores, and negative correlations between glial markers and CR scores, are evident in AD patients. Blue lines indicate linear regression with shaded 95% confidence intervals. Correlation coefficients (R) and significance levels (p-values) are displayed separately within each panel for AD and controls.

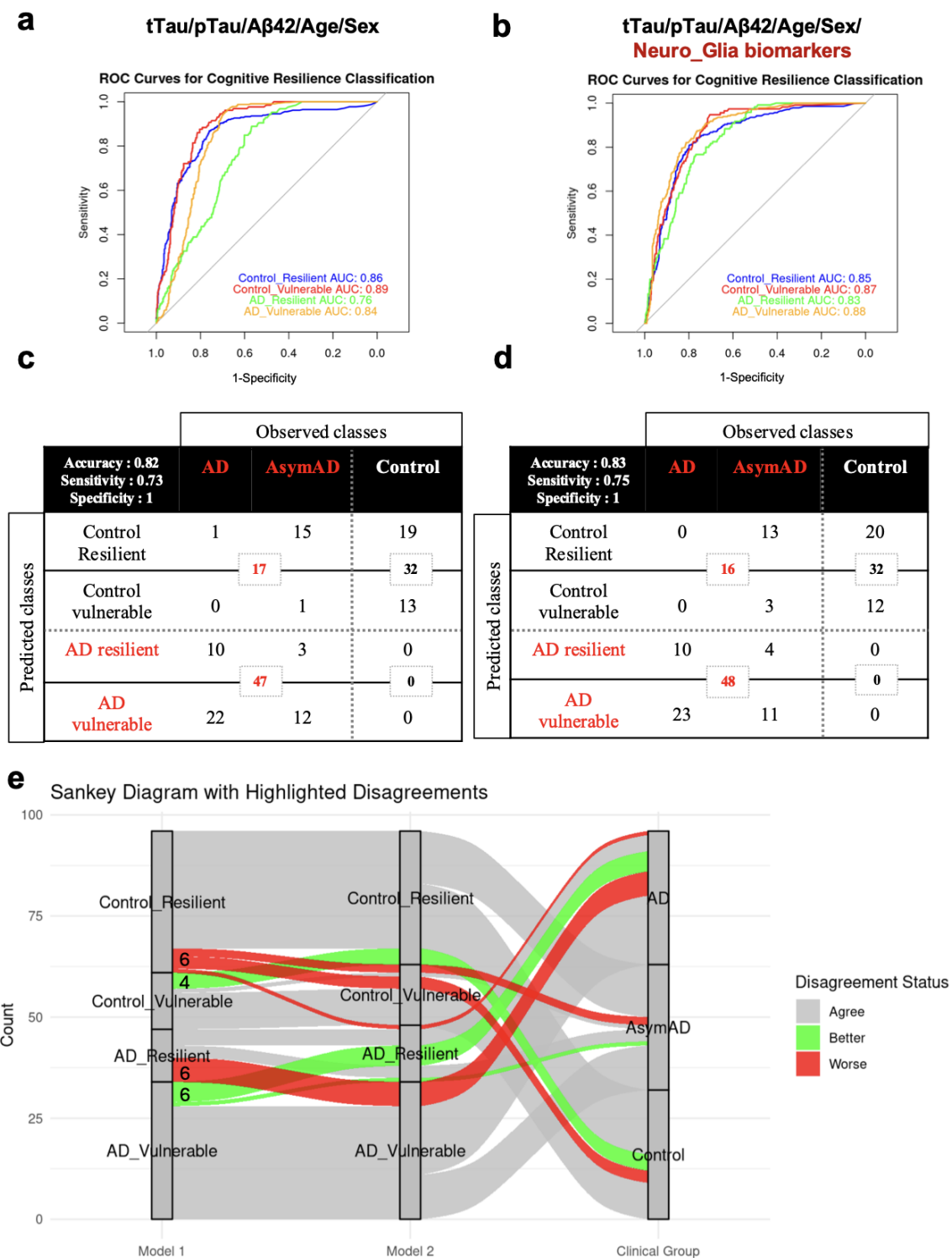


Figure 3-12 Cognitive resilience classification in models with and without neuro-glia biomarkers

a-b) Receiver Operating Characteristic (ROC) curves for the random forest models predicting cognitive resilience classification. Panel (a) shows the baseline model using tTau, pTau, A β 42, age, and sex as predictors, while panel (b) includes the addition of neuro-glia biomarkers (neuronal and glial scores). The ROC curves illustrate improved classification performance

*for AD resilient ($AUC = 0.81$) and AD vulnerable ($AUC = 0.88$) when neuro-glia biomarkers are included, compared to the baseline model ($AUC = 0.73$ for AD resilient, 0.84 for AD vulnerable). Control classifications remained consistent between the two models. **c-d)** Confusion matrices displaying the classification results of the two models on the independent test dataset, consisting of asymptomatic AD (AsymAD), AD, and control individuals. Panel **c)** shows the confusion matrix and performance of the baseline Model 1, and panel **d)** shows Model 2 predictions with neuro-glia biomarkers. Accuracy, specificity, and sensitivity are calculated based on the confusion matrix, where values in the dotted boxes represent the totals for each classification category, summing across AD or control groups irrespective of resilience or vulnerability status. **e)** Sankey diagram highlighting individual level prediction differences in the baseline Model 1 and Model 2, and how they relate to the observed clinical group. Predictions that remain the same between the two models are shown in grey. Predictions with a better (healthier) outcome in Model 2 (Resilient > Vulnerable) are shown in green, and worse (unhealthier) predictions in red.*

3.4 Discussion

Prior large-scale multi-omics analyses have identified complex processes involved in AD pathogenesis^{65,88,90,91,94,98,144} and linked them to clinical and neuropathological traits, including changes during the asymptomatic to symptomatic phase⁷³. My results confirm several of these findings, including the enrichment of AD risk loci in microglia, consistent identification of established AD pathways, evidence for heterogeneous cognitive decline trajectories, and validation of previously proposed CSF and plasma biomarkers. My approach also offers important novel perspectives. First, the use of joint transcriptomics-proteomics modules allows precise identification of glial mechanisms underlying AD heterogeneity—extending beyond existing bulk or single-omic frameworks. Second, through pseudotemporal ordering of postmortem samples from different brain regions, I infer a temporal sequence of glial responses: transcriptomic microglial activation (linked to AD genetic risk) emerges earlier in the disease course, while proteomic astrocytic responses become prominent later and are associated with accelerated cognitive decline. I show that the neurotoxic astrocyte phenotype is detectable in human proteomic data and correlates with synaptic dysfunction (e.g., impaired SNAP25-VAMP interactions) in individuals experiencing rapid decline. Finally, to situate my findings within the broader landscape of AD research, I related my results to a published CSF/plasma proteomic study by Dammer and colleagues and showed the potential for using neuroglia biomarkers for patient stratification in AD.

3.4.1 Glial activation dynamics

My unique contribution lies in identifying specific transcriptomics-proteomics modules (F14_UP_ME1, F14_UP_ME2, F11_DOWN_ME2) representing microglia, astrocytes, and neuronal injury mechanisms. I leveraged an approach to study cross-sectional *post mortem* data through pseudotime analysis, which leverages unsupervised integration of different brain regions (MTG, SOM) affected at various disease stages. I build on previous single-cell studies identifying glial subpopulations involved in AD^{29,40}, validating the cell-state specificity of AD-correlated glial activation and clarifying which subpopulations may mediate AD genetic risk. Subpopulations such as Micro2, enriched in HLA/DAM, and Astro3&6, enriched in neurotoxic reactive and exhausted astrocyte genes, were characterised. Contrary to findings suggesting a protective role for the microglia HLA state²⁹, this data indicate otherwise, possibly due to the mixed DAM/HLA composition in the Micro2 subpopulation. This mixed population aligns with the initial pseudotime trajectory analysis from Mancuso et al., with an evolution from DAM to HLA. I emphasised the distinct contributions of transcriptomics and proteomics in mediating AD genetic risk in microglia and astrocytes, with transcriptomics revealing numerous transcription factors regulating the microglial network (CEBPB, STAT1), while the neurotoxic mixed reactive/exhausted astrocyte network may predominantly be regulated post-transcriptionally.

Previous research shows AD PRS is linked to cognitive decline rates, mainly through A β pathology¹⁸⁷. My findings align with this, demonstrating direct and indirect effects of pathway-specific PRS on AD progression. Indeed, results suggest that genetic risk is a determinant of microglial activation in AD, corroborating previous research^{69,188}. By contrast, I hypothesise that a neurotoxic mixed reactive/exhausted astrocyte response, which appear to have a lower genetic association, is related to local environmental cellular and neuropathology signals. This aligns with iPSC model findings showing *TREM2*^{R47H} promotes proinflammatory microglia³⁰, while *APOE*- ϵ 4 increases A β aggregates and tau hyperphosphorylation, with *APOE*- ϵ 4 astrocytes showing impaired A β uptake and cholesterol accumulation¹⁸⁹. Moreover, Serrano-Pozo et al. were recently the first group to propose that an additional “exhausted” astrocyte state might deviate from the traditional homeostatic-to-reactive trajectory, reflecting a maladaptive response driven by a prolonged exposure to neuropathology⁴⁰. Notably, metallothionein, lipid and inflammatory signalling, and extracellular matrix-related pathways,

a hallmark of these exhausted astrocytes, were also characteristic of the F14_UP_ME2 signature (MT1F, TGFB1, PLEC).

3.4.2 Differential cognitive decline groups and their molecular correlates in early AD

Differential glial activation may also partly explain some of the heterogeneity in clinical outcomes, including variations in the rate of disease progression, such as fast or slow progression, as recently suggested⁴⁹. I focused the differential cognitive decline analysis on Braak stages III-IV, which is considered a suitable therapeutic window for disease-modifying treatments as this phase marks the onset of clinical symptoms¹⁹⁰. Via multi-omics clustering of transcriptomics and proteomics, I identified two early AD molecular subgroups within these stages. These were subsequently annotated using retrospective assessment of lifetime cognitive trajectories and found to be related to fast and slow cognitive decline groups. This aligns with previous studies that have uncovered heterogeneous cognitive trajectories in mild cognitive impairment⁶³. In another study of individuals at Braak III-IV stages, changes in glial cell phenotypes were shown to occur before severe neurofibrillary tangle formation and correlated with markers of cortical cell damage, but more importantly, could differentiate symptomatic from resilient AD¹⁹. I linked this accelerated decline to an increased neurotoxic mixed reactive/exhausted astrocytic response and neuronal injury response predominantly found in RORB⁺ excitatory neurons, suggesting a particular susceptibility corroborating previous research^{185,191}.

I then sought to link this differential cognitive decline to synaptopathy. Synaptic changes and SNARE transcriptional reduction are associated with cognitive deficits¹⁹² and the concept of "cognitive reserve", where higher baseline synaptic functionality helps resist neuropathological insults¹⁸⁰. Although baseline differences were not assessed *post mortem*, I observed a steeper relationship between the rate of cognitive decline and SNAP25-VAMP synaptic functionality in the fast-progressing group, which may be influenced by biological vulnerability, synaptic reserve, plasticity, and accelerated pathological processes. SNAP25-VAMP synaptic functionality was associated with the neuronal injury response module, F11_DOWN_ME2, highlighting variation related to synaptic dysfunction. Given that astrocytes are essential for

synaptic plasticity and maintenance, I propose that the neurotoxic mixed reactive/exhausted phenotype may exacerbate synaptic dysfunction involving SNAP25 and VAMP proteins. This raises the hypothesis that accelerated cognitive decline may be driven by neurotoxic astrocytes undergoing a progressive transition from intermediate reactive to exhausted states and warrants further investigation into this maladaptive trajectory. Astrocyte contribution to synaptopathy could be through a direct neurotoxic effect, as well as an indirect impact resulting from the loss of normal astrocytic support at the synapse, thereby linking the exhausted phenotype to characteristics of “neuro-neglect”¹⁹³. Also, the higher neuroprotective response might indicate a compensatory mechanism, aligning with findings that link Clusterin (CLU) from astrocytes to rescue synaptic deficits in AD models¹⁹⁴. This also fits in with GWAS studies that identified a protective role for SNPs in the CLU locus¹⁹⁵, and these were recently mechanistically linked to beneficial neuro-astrocyte communication¹⁹⁶.

3.4.3 Identification of potential CSF and plasma biomarkers for rapid cognitive decline

My study also establishes the translational relevance of these findings by identifying overlapping molecular signatures in CSF and plasma, which could serve as biomarkers for identifying individuals at risk of faster cognitive decline. Notably, sex was accounted for in the analysis, as disease-associated modules in the discovery cohort were identified based on differential module eigengene expression while adjusting for sex as a covariate. Many of these biomarkers corroborate previous research in other larger studies, including SNAP25¹⁹⁷, neurofilament light chain¹⁹⁸ (NFL), GAP43¹⁹⁹, or complement proteins²⁰⁰. Furthermore, other proteins identified within these modules, such as Visinin-like 1 (VSNL1) or β -synuclein (SNCB), could represent novel biomarkers and warrant further investigation. Previous studies have shown that a synapse protein signature (YWHAG-to-NPTX2 ratio) in cerebrospinal fluid and plasma predicts cognitive maintenance versus decline in AD⁵⁴. Therefore, the identified candidates were tested in CSF to predict resilience or vulnerability status, defined as individuals exhibiting more or less cognitive impairment than expected given the extent of AD pathology. The baseline models had AUCs ranging between 0.76 to 0.89, comparable to what has been recorded in the literature ranging from 0.81¹⁶ to 0.9²⁰¹. The neuro-glia score combining GAP43, NPY, C3, CD14, and AZGP1 proteins successfully improved the

identification of vulnerable or resilient AD compared to predictions using A β 42, tTau, pTau and age alone (Improvement of AUC from 0.84 to 0.88). In other words, the candidate biomarkers initially identified to be linked to the rate of cognitive decline in early AD were also shown to improve prediction of cognitive resilience or vulnerability. In the early AD cohort, the cognitive vulnerability was associated with the neurotoxic mixed reactive/exhausted astrocyte and synaptic module activity but not activated microglia, which suggests that the interaction of the two, possibly at the tripartite synapse, plays a more direct role in vulnerability. Overall, I empirically demonstrated that the rate of cognitive decline is indeed linked to the degree of cognitive resilience or vulnerability, which the proposed markers can predict.

3.4.4 Study limitations

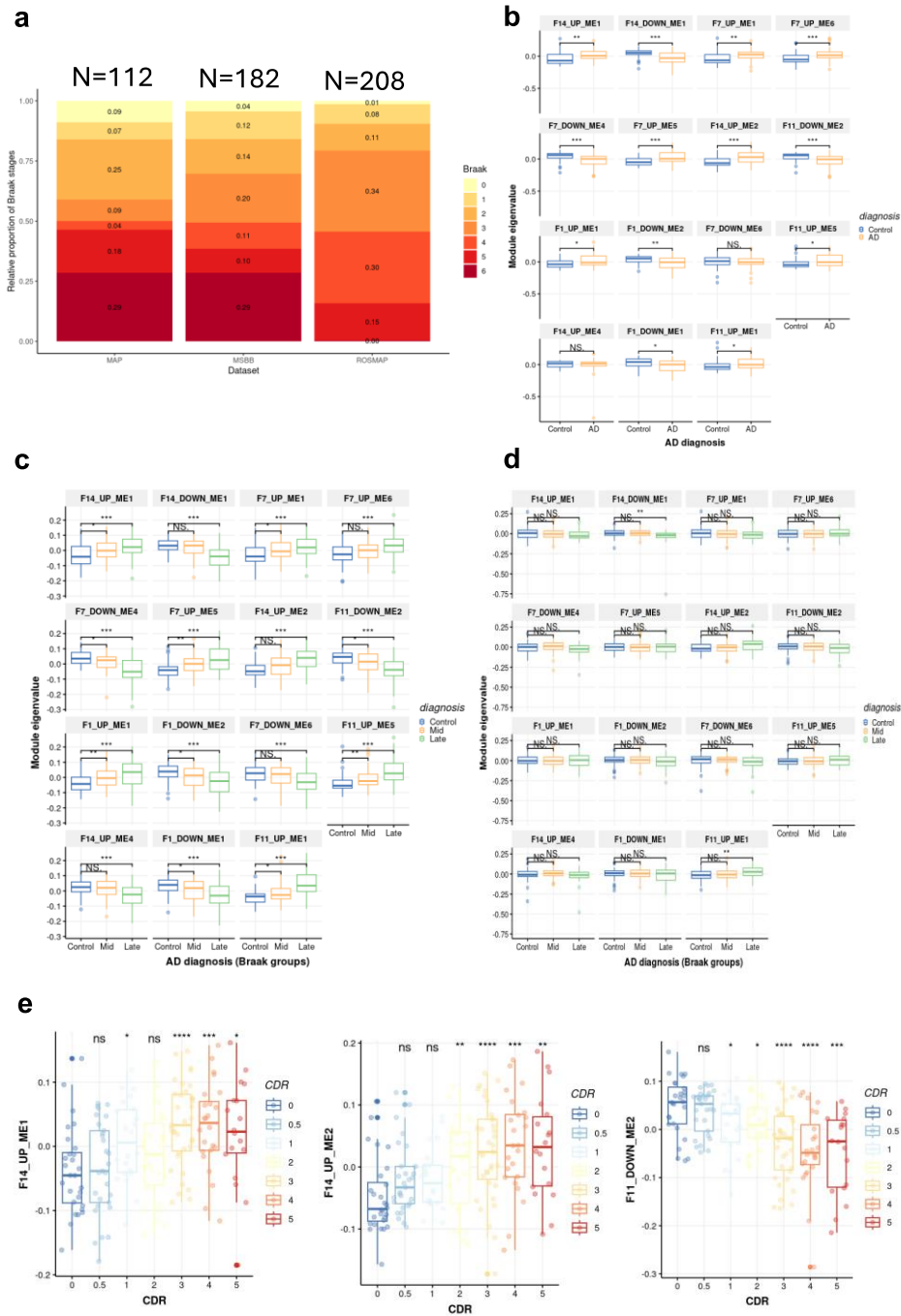
The strengths of my study relied mainly on the comprehensive multi-regional, multi-omics, multi-resolution set-up. A limitation of this study is the limited proteome depth (~3,000 proteins) compared to transcriptomics (~20,000 transcripts), which likely resulted in missing proteomic changes from the internal cohort. Similarly, the removal of the least explanatory transcripts to achieve a balanced MOFA integration may have led to missing important transcripts. Altogether, the MOFA factors and derived AD-correlated modules may have overlooked critical genes/proteins and pathways involved in the disease process. Furthermore, the initial bulk cohort included comprehensive bulk transcriptomic and proteomic measurements in the MTG and SOM per donor. However, the follow-up analyses, including snRNA-seq and histopathology, were completed on a subset of samples due to tissue quality and availability. This means that not all bulk proteomic/transcriptomic donors were fully represented in the follow-up analyses, so correlations and insights derived from the integrated data may be less representative of the whole cohort. The sample size also remains a limitation. The discovery cohort consisted of only 56 individuals with complete bulk omics data for disease module inference (33 AD; 23 Controls), while the external ROSMAP validation cohort included 132 individuals. This modest cohort size may limit statistical power, particularly for detecting subtle effects or rare cell states, and may reduce the generalisability of some findings.

Another limitation of this study is the analysis of cross-sectional post-mortem data via pseudotime, which represents snapshots rather than continuous progression. To mitigate this and ensure the observations reflect dynamic disease processes, I related AD-correlated modules to longitudinal cognitive data from an independent cohort in early AD. To address potential survivor bias, I restricted the subset to those with follow-ups of 12 years or less. Furthermore, differences in population representation (ROSMAP early AD cohort mean of 89 years old vs. US average 83.7 years old²⁰²) limit the generalisability of the findings. Finally, the analysis to test the candidate CSF markers may have been limited in accuracy because not all candidates were available in the dataset. I also did not use any plasma markers where fewer candidates were flagged in the module overlap analysis, further limiting the translational value of my findings.

3.5 Conclusions and future work

This integrative multi-omics study connects glial activation to synaptopathy and cognitive decline in AD. Future research should apply integrative omics approaches prospectively and longitudinally, utilising CSF and plasma biomarkers to confirm glial activation dynamics and their effects on cognitive decline. The identified CSF and plasma biomarkers hold significant potential for stratifying rapidly progressing cases and identifying disease trajectories, warranting further exploration. As a next step, I will focus on analysing the follow-up cognitive assessments for patients that the model predicted to be either AD-resilient or AD-vulnerable. This will allow us to validate whether baseline CSF markers can predict future clinical progression in living patients. I anticipate the ones predicted to be vulnerable to follow a fast progression trajectory, and vice versa. Synaptic multi-omics measurements will enable the investigation of the mechanisms involving neurotoxic astrocytes and their potential role in SNARE synaptic dysfunction, providing insights into how these processes might accelerate synapse loss. *Chapter 4* will explore the analysis of synaptic proteomics to examine the potential connections between reactive neurotoxic astrocyte processes at the tripartite synapse and their impact on synaptopathy.

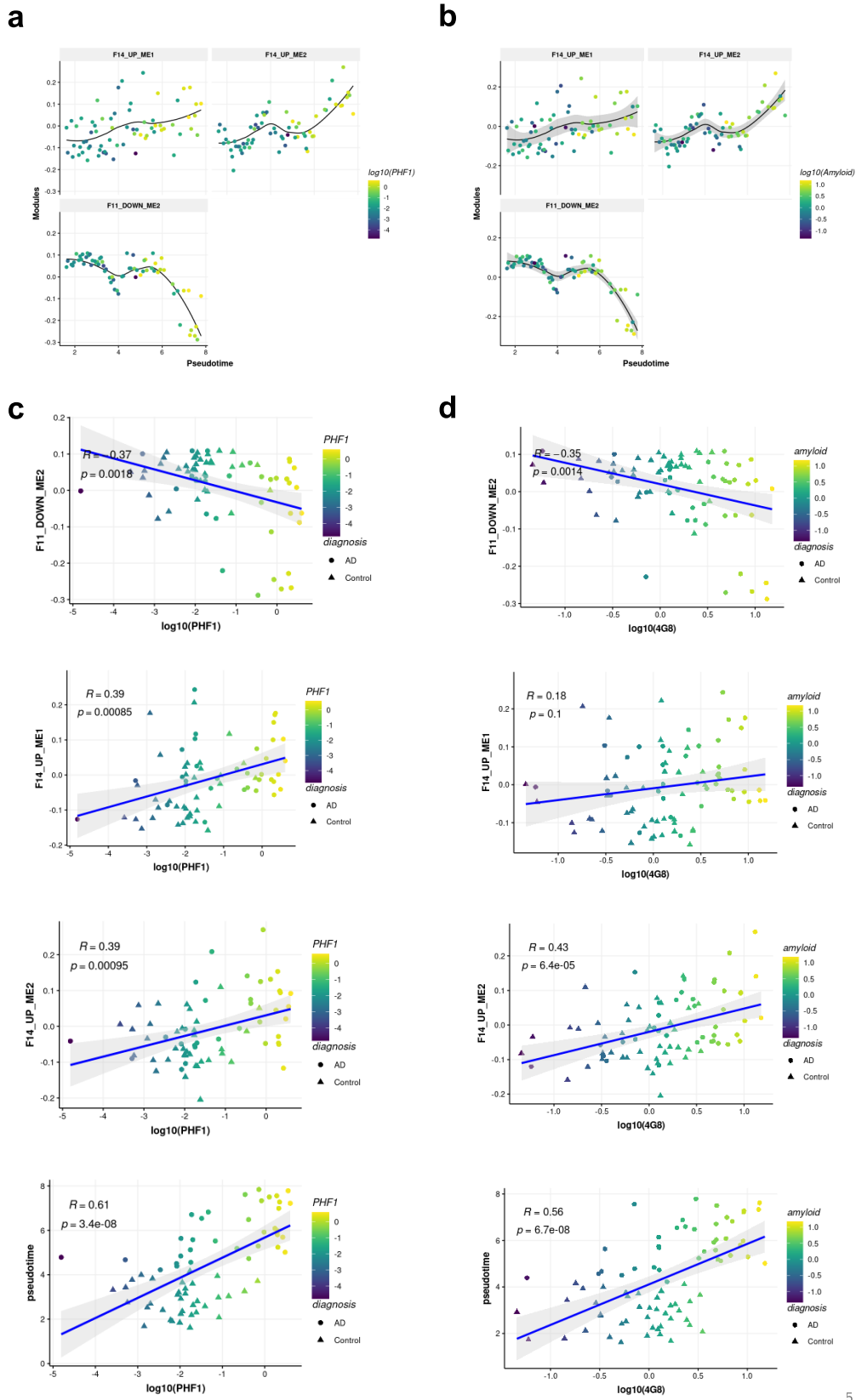
3.6 Supplementary figures



Supplementary figure 3-1 Module eigenvalue validation in external cohorts (ROSMAP, MSBB)

a Discovery cohort (MAP) and validation cohorts (MSBB, ROSMAP) repartition of samples by Braak stages. MSBB and MAP show a similar Braak distribution, while ROSMAP has a higher representation of Mid Braak (III-IV). **b** Module differences in the discovery cohort. **c** Module differences in MSBB, demonstrating consistent patterns of up-regulation and down-regulation of module eigenvalues in AD compared to controls, similar to those observed in the discovery cohort. **d** Module differences in ROSMAP do not show significant differences, which

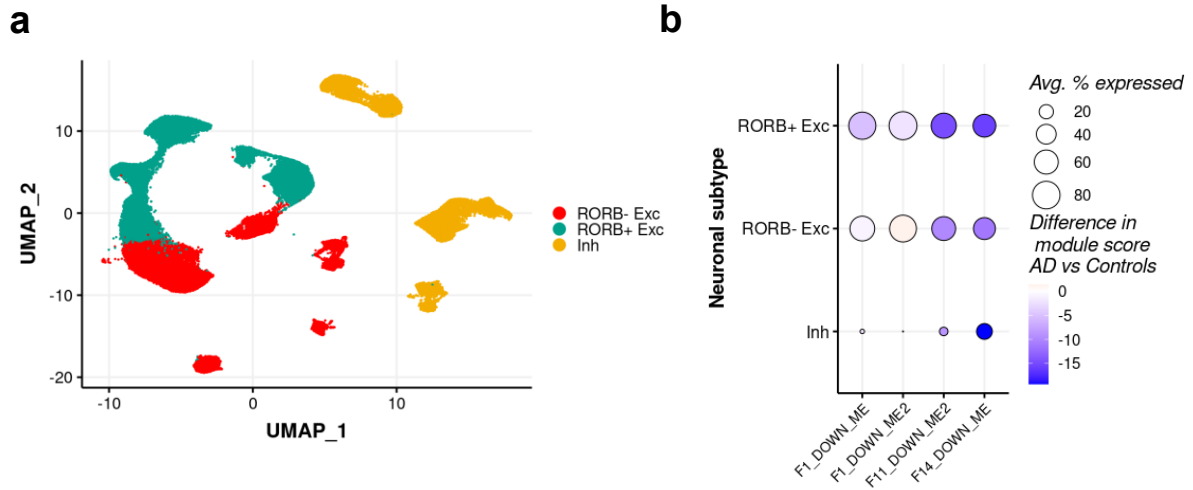
is attributed to the cohort composition (more middle stage and fewer early and late-stage cases). e) CDR differences in microglia, astrocyte and neuronal injury modules in MSBB. (Two-sided Wilcoxon rank-sum test, adjusted p , $*p \leq 0.05$, $**p \leq 0.01$, $***p \leq 0.001$).



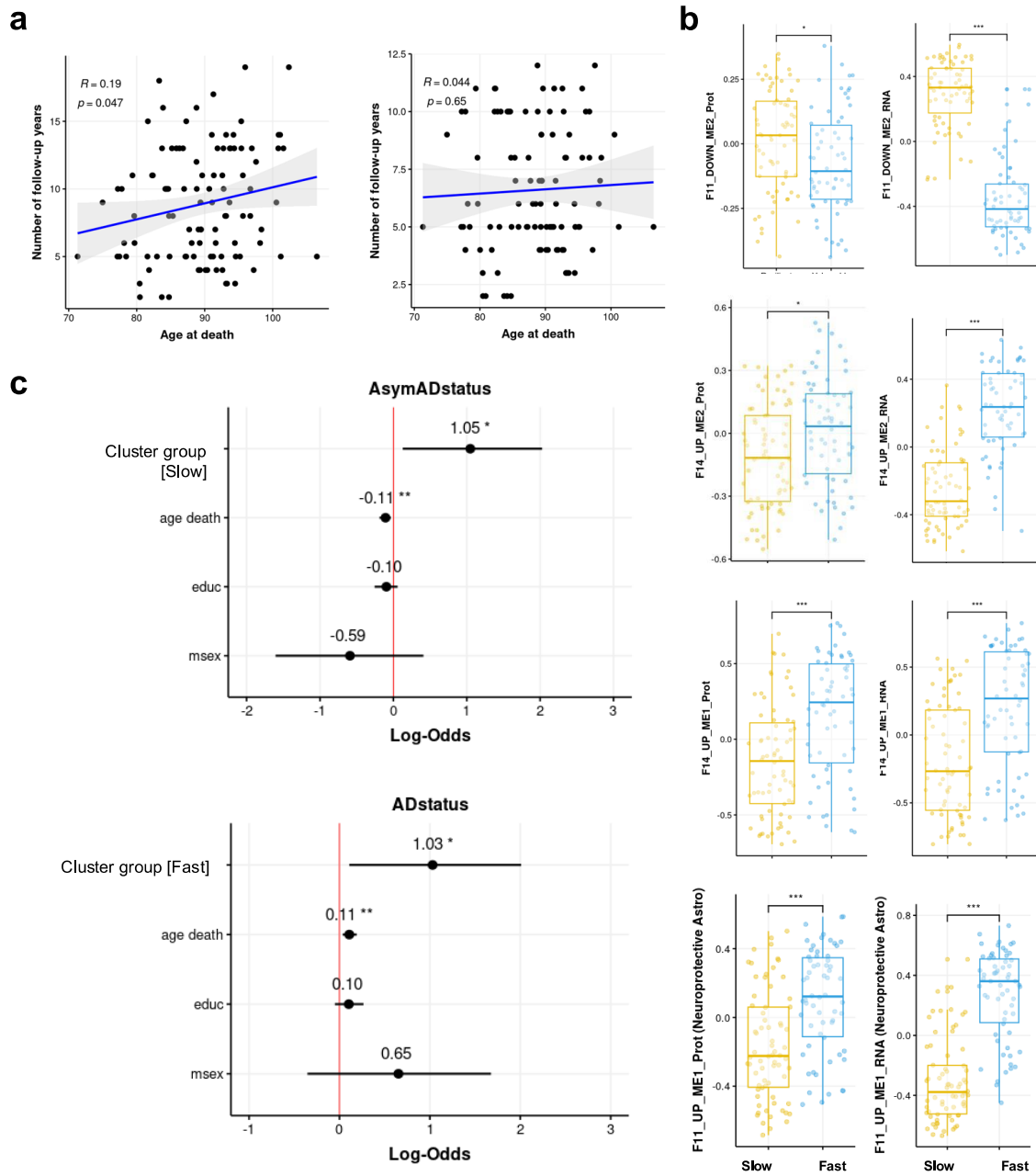
5

Supplementary figure 3-2 Module trajectories alongside pseudotime and association to neuropathology

a) Module eigenvalue as a function of pseudotime (loess regression). Points coloured by $\log_{10}(\text{PHF1})$ **b)** Same as **a)** but points are coloured by $\log_{10}(\text{amyloid})$ load. **c)** Synaptic module, microglia, astrocyte module eigenvalue and pseudotime as a function of $\log_{10}(\text{PHF1})$ **d)** Same as in **c)** but as a function of $\log_{10}(\text{amyloid})$ measured by 4G8.

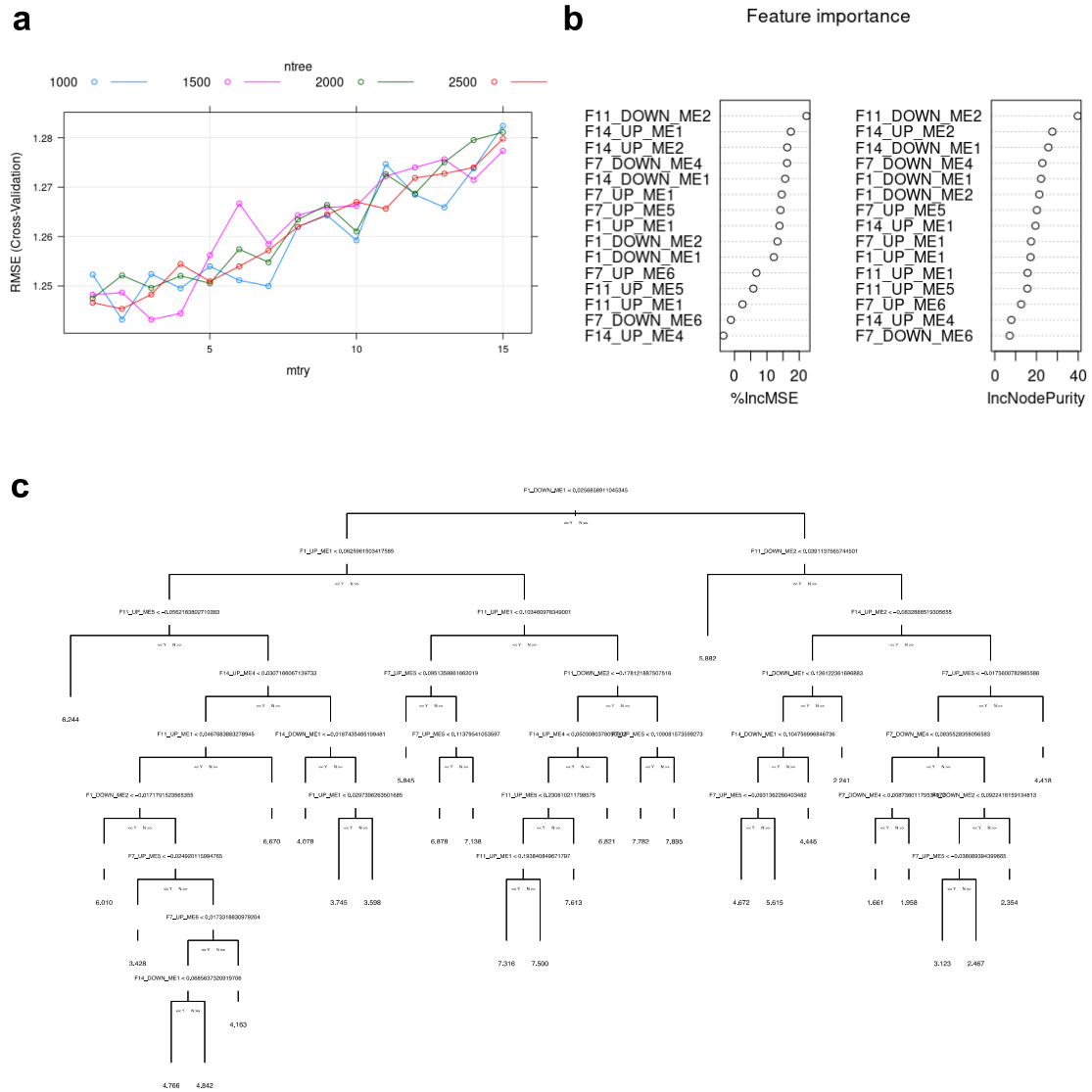


Supplementary figure 3-3 Neuronal subpopulations and neuronal bulk module projections
a) UMAP visualisation of neuronal subtypes in snRNA-seq data grouped by RORB positivity in excitatory subtypes. **b)** Projecting AD-related neuronal modules on single-cell clusters using HdWGCNA grouped by RORB positivity in excitatory subtypes.



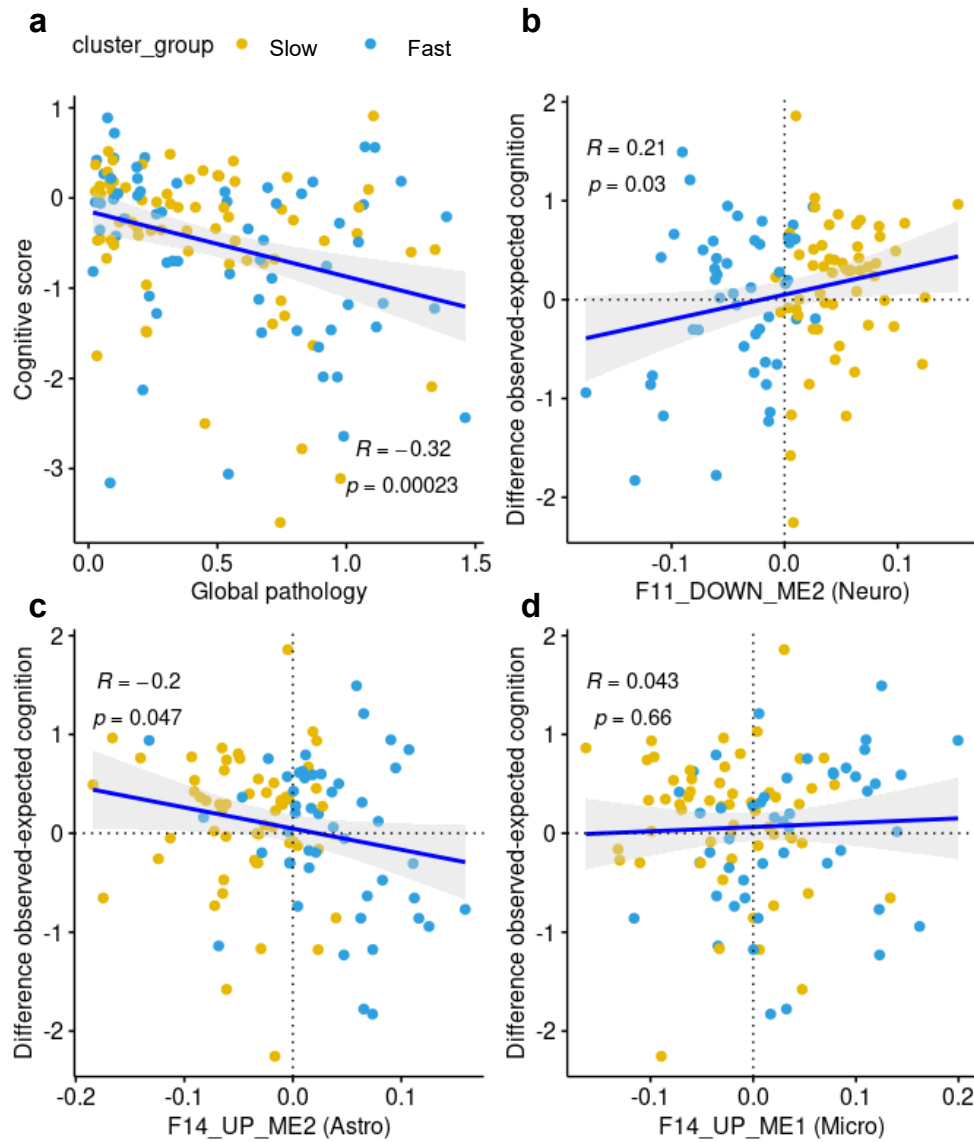
Supplementary figure 3-4 ROSMAP cohort in early AD – multi-omics clustering

a) Relationship between number of follow-up years as a function of age at death to detect survivor bias in early AD ROSMAP cohort ($n=132$). Bias is corrected for when filtering for participants with less than 12 years of follow-up ($n=102$). **b)** Module eigenvalue differences in different modules are separated by proteomics and transcriptomics. (Two-sided Wilcoxon rank-sum test, adjusted p , * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$) **c)** Logistic regression analysis assessing the influence of cluster group on AD symptomatic ($MMSE < 25$) or asymptomatic ($MMSE \geq 25$) status, adjusted for age at death, education level, and sex. Odds = $e^{\log\text{-odds}}$. $e^{1.05} \approx 2.85$; $e^{1.03} \approx 2.80$.



Supplementary figure 3-5 Random Forest model to predict pseudotime

a) Hyperparameter tuning results for the random forest model. The graph shows the RMSE values obtained through 10-fold cross-validation, for different values of *mtry* (number of variables randomly sampled as candidates at each split) and *ntree* (number of trees in the forest). The coloured lines represent different *ntree* values (1000, 1500, 2000, 2500). The optimal hyperparameters were identified as *mtry* = 3 and *ntree* = 1500, indicated by the lowest RMSE values. **b)** Variable importance plots for the random forest model. The left plot shows the percent increase in mean squared error (%IncMSE) when each variable is permuted, indicating the impact of each variable on model accuracy. The right plot shows the increase in node purity (IncNodePurity) for each variable, representing their contribution to reducing impurity in the decision trees. Both plots highlight the most influential features used in predicting pseudotime. **c)** Tree illustrating the decision rules and splits made based on different variables and their thresholds. The tree nodes display the variable names, split points, and the predicted pseudotime values at the terminal nodes (leaves).



Supplementary figure 3-6 Disease modules and cognitive resilience in early AD ROSMAP

a) Linear regression plots showing the relationship between cognitive scores and the global pathology level. Fitted values on the line correspond to the expected cognition based on global pathology. **b)** Lower values of F11_DOWN_ME2, indicating more synapse dysfunction, are associated with worse-than-expected cognitive performance, indicating that this module is associated with cognitive vulnerability. **c)** higher F14_UP_ME2 (Astro) values are associated with worse-than-expected cognitive performance, indicating that this module is associated with cognitive vulnerability. **d)** No association between microglial module and cognitive resilience.

3.7 Supplementary tables

https://eleonoreschneegans.shinyapps.io/supplementary_3/

Supplementary Table 3-1	Discovery of cohort clinical characteristics
Supplementary Table 3-2	Bulk omics processing parameters
Supplementary Table 3-3	MOFA variance explained per factor
Supplementary Table 3-4	Module membership
Supplementary Table 3-5	Module Pearson's correlation to neuropathology & pseudotime
Supplementary Table 3-6	Module differential expression against pseudotime
Supplementary Table 3-7	Module differential expression against AD vs. Controls
Supplementary Table 3-8	Cameron et al. (2024) neuroprotective and neurotoxic genes
Supplementary Table 3-9	MAGMA SNP to Gene (AD GWAS Meta analysis genes.out)
Supplementary Table 3-10	Microglia PRS
Supplementary Table 3-11	Astrocyte PRS
Supplementary Table 3-12	iCluster early AD posterior probability
Supplementary Table 3-13	early AD subgroups clinical characteristics
Supplementary Table 3-14	Longitudinal cognitive measurements in subgroups (interaction model)
Supplementary Table 3-15	Early AD clusters differential expression Fast vs. Slow (Proteomics/ Transcriptomics)
Supplementary Table 3-16	Presynaptic proteins and SNARE protein-protein interactions in early AD subgroups
Supplementary Table 3-17	Neuronal injury response module ~ cognition + cov
Supplementary Table 3-18	Rate of decline ~ Synaptic functionality * cluster group + cov
Supplementary Table 3-19	Synaptic functionality ~ astro* cluster group + cov
Supplementary Table 3-20	Synaptic functionality ~ micro* cluster group + cov
Supplementary Table 3-21	Differential expression astro (amyloid)
Supplementary Table 3-22	Differential expression astro (PHF1)
Supplementary Table 3-23	Differential expression micro (amyloid)
Supplementary Table 3-24	Differential expression micro (PHF1)

3.8 Supplementary files

Supplementary file 1	QC Report	QC.html
Supplementary file 2	Microglia module Pathway Report	F14_UP_1.html
Supplementary file 3	Neurotoxic Astrocyte module Pathway Report	F14_UP_2.html
Supplementary file 4	Neuroprotective Astrocyte module Pathway Report	F11_UP_1.html
Supplementary file 5	Synaptic module Pathway report	F11_DOWN_2.html
Supplementary file 6	Proteins up-regulated in fast cluster Pathway report	proteomics_fast_UP.html
Supplementary file 7	Genes up-regulated in fast cluster Pathway report	transcriptomics_fast_UP.html
Supplementary file 8	Proteins down-regulated in fast cluster Pathway report	proteomics_fast_DOWN.html
Supplementary file 9	Genes down-regulated in fast cluster Pathway report	transcriptomics_fast_DOWN.html

Chapter 4 A synaptic-astrocytic proteomic signature associated with synaptopathy in Alzheimer's disease

Chapter declarations:

1. Some of the work in this chapter has been previously made available as a preprint:

Griffiths Jessica[#], El onore Schneegans[#], Harry Whitwell, Zhen Qiu, Beverley Notman, Dorcas Cheung, Nanet Willumsen, Paul M Matthews, Seth G N Grant and Johanna S Jackson (January 2025). "A synaptic-astrocytic proteomic signature associated with synaptopathy in Alzheimer's Disease". In: *bioRxiv*. doi: <https://doi.org/10.1101/2025.01.23.634408>

[#] These authors contributed equally

[CC-BY-NC-ND 4.0 International license](#)

2. I thank the following contributors:

All the internal data used for analysis was generated entirely by the wet lab team, specifically:

- Jessica Griffiths: Syn-PER and Synaptome mapping.
- Harry Whitwell: Label-free quantitative proteomics.
- Johanna Jackson: Synaptoneurosome staining.

Abstract

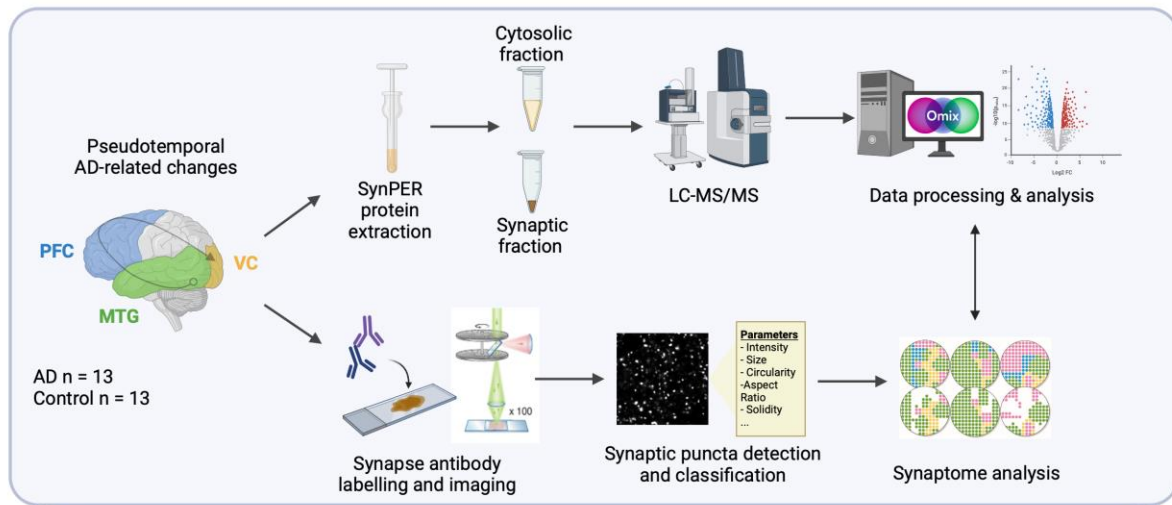
Background: Synapse loss is the strongest neuropathological correlate of cognitive impairment in Alzheimer's disease (AD). Therefore, the synapse is a critical substrate linking cognitive symptoms and proteinopathy in AD, making it a promising target for novel therapies. Astrocytes form the tripartite synapse and contribute to synapse formation, maintenance, and elimination — making them potential players in synaptopathy in AD.

Methods: We used synapse proteomics and synaptome mapping techniques to define a spatiotemporal profile of synaptic protein alterations and analyse NMDA receptor subunit GluN1⁺ excitatory synapse loss in late-stage AD and non-diseased control *post mortem* samples from three brain regions, each representing different stages of AD pathology progression. By integrating proteomics data from synaptic and cytosolic fractions, I further examined the autonomy of synapses and their relationship within the broader cellular environment in AD.

Results: We characterised a synaptic proteomic cascade of events leading to synapse loss alongside an associated synaptotoxic astrocyte proteomic signature. Briefly, we characterised regional molecular pathology at the synapse level and revealed differential excitatory synapse subtype vulnerability in AD vs. controls. The loss of mature GluN1⁺ excitatory synapse subtypes was intricately associated with astrocytic proteins enriched in peri-synaptic astrocytic processes (PAPs). These PAPs were enriched in proteins related to complement activation and endocytosis, pathways implicated in neuroinflammatory responses and synaptic pruning, further suggesting a neurotoxic astrocyte phenotype localised to the tripartite synapse in AD.

Conclusion: This study provides novel insights into the pseudotime course, regional specificity, and molecular pathways underpinning synapse vulnerability in AD. The findings highlight astrocyte reactivity to be associated with synaptopathy, and underscore PAPs as potential therapeutic targets for preserving synaptic resilience and alleviate cognitive decline in AD.

Graphical abstract



Proteomics integration of synaptic and cytosolic fractions

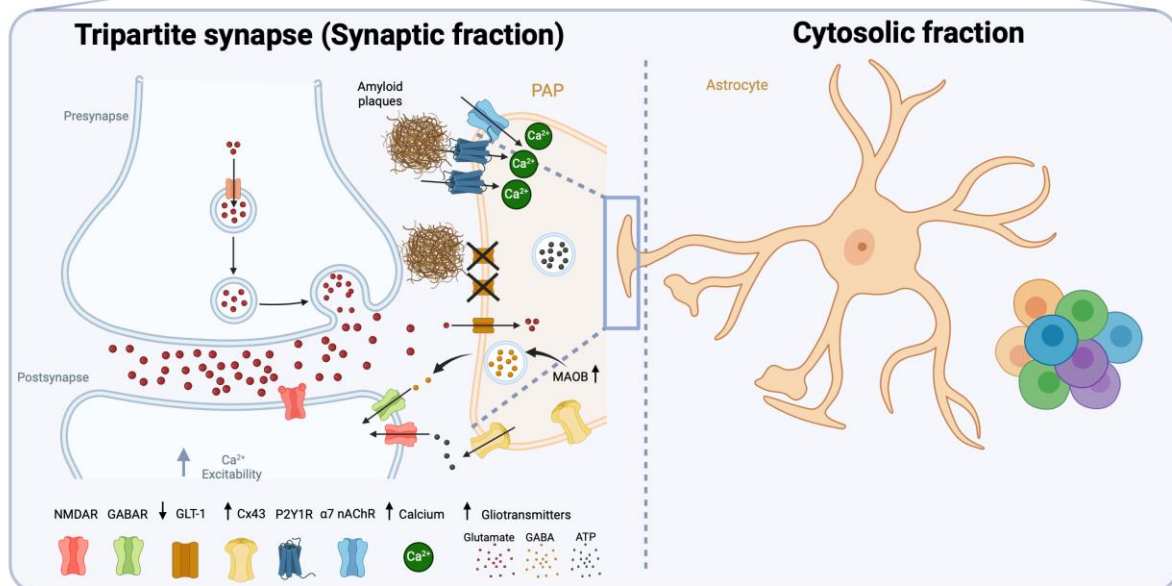


Figure 4-1 Chapter 4 Graphical abstract

Figure generated via [Biorender](#).

4.1 Introduction

4.1.1 Studying synaptopathy in AD using proteomics

Synaptopathy is defined as a state associated with the dysfunction, structural alteration, or loss of synapses, which impairs neural communication and contributes to the progression of neurodegenerative diseases such as AD. Synapse dysfunction and loss are recognised as some of the earliest hallmarks of AD^{50–52}, occurring decades before the onset of clinical symptoms and closely correlates with the cognitive impairment associated with the disease^{53–56}. In other words, synapse loss is not merely a result of neuronal loss since the remaining neurons exhibit fewer dendritic spines in AD^{203,204}. *Post mortem* human brain AD studies reveal region-specific patterns of synapse loss, depletion of synaptic proteins, and distinct morphological alterations, altogether pointing to selective vulnerability of specific synaptic subtypes²⁰⁵. Studies in preclinical AD models and human AD showed a differential susceptibility between pre- and post-synaptic components^{206,207}. Consequently, the synapse is a critical substrate linking cognitive symptoms and proteinopathy in AD, making it a promising target for novel therapies^{208,209}.

Synaptic proteins play essential roles in signal transduction and plasticity, fundamental to synaptic function. These proteins are synthesised in the soma before being trafficked to synapses or locally at the synapse from trafficked mRNA²¹⁰. Synaptoneurosomes, subcellular fractions containing both pre- and post-synaptic components²¹¹, have been extensively used in studies of synaptic dysfunction in AD, primarily for targeted proteomic analyses such as Western blotting. More recently, omics approaches have been applied to analyse these fractions, particularly in experimental models, although their use in human studies remains limited. Thermo Scientific's Synaptic Protein Extraction Reagent (Syn-PERTM) method enriches synaptoneurosomes while preserving functionally active synaptic proteins and allows the interrogation of their proteome. This method retains non-neuronal synaptic components²¹¹, such as peri-synaptic astrocytic processes (PAPs).

4.1.2 Reactive astrocytes, PAPs and synaptopathy in AD

One astrocyte can connect with as many as 2,000,000 synapses²¹². Astrocytes can release gliotransmitters like ATP and glutamate, thereby acting as a partner in synaptic signalling²¹³. Additionally, astrocytes express numerous synaptic genes, shared with neurons, and are involved in critical processes such as vesicular transport, synaptic vesicle exocytosis, and synaptogenesis²¹⁴. Peri-synaptic astrocytic processes (PAPs) are now widely acknowledged as part of the “tripartite synapse”²¹⁵. PAPs are fine (<50nm), branched extensions of astrocytes surrounding synapses, that play an important role in synaptic homeostasis, synaptogenesis, and pruning^{216,217}.

AD induces astrocyte morphological and functional changes, leading to reactive astrogliosis, which has been extensively characterised through snRNA-seq studies²¹⁸. Imaging studies using PET tracers have suggested that astrogliosis is an early event in AD, with some clinical studies demonstrating elevated astrocyte reactivity as early as 20 years before the onset of clinical symptoms^{219,220} and has also been significantly correlated with cognitive function in AD patients²²¹. Transcriptomics and proteomics studies highlight the heterogeneity of reactive astrocyte markers, with no consensus on a defining signature, as demonstrated by works from Liddel³², Habib³⁷, Sadick²²², and others^{40,70,162,223–226}. The role of reactive astrocytes in AD-related synaptopathy has only recently begun to emerge^{27,227,228}. The interplay between astrocytes and synapses is complex, and reactive astrogliosis in AD can lead to synaptic dysfunction and loss through multiple mechanisms, summarised below and visualised in *Figure 4-2*.

1. **Glutamate excitotoxicity:** Reactive astrocytes surrounding A β plaques downregulate the glutamate transporter GLT-1, reducing their capacity to clear glutamate from the synaptic cleft and thus prompting its accumulation and synaptic damage²²⁹. Cognitively normal patients exhibiting AD pathology show stable GLT-1 expression, suggesting its role in preserving cognitive functions potentially via mechanisms of synaptic resilience²³⁰.
2. **Pathological calcium dysregulation:** A β plaques react with P2Y1R or α 7 nAChRs, which leads to increased intracellular Ca²⁺ in astrocytes²³¹. Elevated calcium levels activate the Ca²⁺-dependent phosphatase calcineurin and calcineurin/NFAT signalling pathway and drive neuroinflammation and hyperexcitability²³². Intracellular calcium elevation has been

linked to the dysregulated release of gliotransmitters ATP and glutamate via the Cx43 hemichannel (GJA1 gene)²³³ or vesicle release, which alters synaptic modulation^{234,235}.

3. **Oxidative stress:** Astrocytes express MAOB, an enzyme that mediates GABA synthesis and contributes to tonic inhibition. In AD, elevated MAOB levels lead to aberrant GABA synthesis in astrocytes, resulting in excessive suppression of neuronal activity and further oxidative stress, exacerbating neuronal degeneration and astrocytic reactivity^{236–238}. Pharmacological inhibition of MAOB has been shown to attenuate GABA levels, reduce astrogliosis, enhance synaptic transmission, and rescue cognitive impairments in AD mouse models²³⁹. Simultaneously, during the process of MAOB-mediated GABA synthesis in astrocytes, H₂O₂ is generated as a byproduct, inducing oxidative stress that critically contributes to the degeneration of the neighbouring neurons and further exacerbates astrocytic reactivity²⁴⁰.
4. **Secretion of toxic molecules:** Diverse neurotoxins including LCN2²⁴¹ or proinflammatory chemokines such as CXCL10, CXCL12, CXCL2 can be secreted from astrocytes²⁴².
5. **Aberrant synaptic pruning:** A pathological re-introduction of mechanisms used during development to eliminate weaker synapses may lead to removing functional and degenerating synapses in AD. This pruning may be mediated by "eat-me" signals, including complement components (C1q, C3) and milk fat globule-EGF factor 8 (MFG8)⁴⁴, which are expressed on synaptic membranes and can trigger astrocytic engulfment of synapses²²⁷.
6. **Dysfunctional water and ion homeostasis:** Altered AQP4 expression has been linked to impaired glymphatic clearance, leading to the accumulation of toxic metabolites or A β around synapses²⁴³.

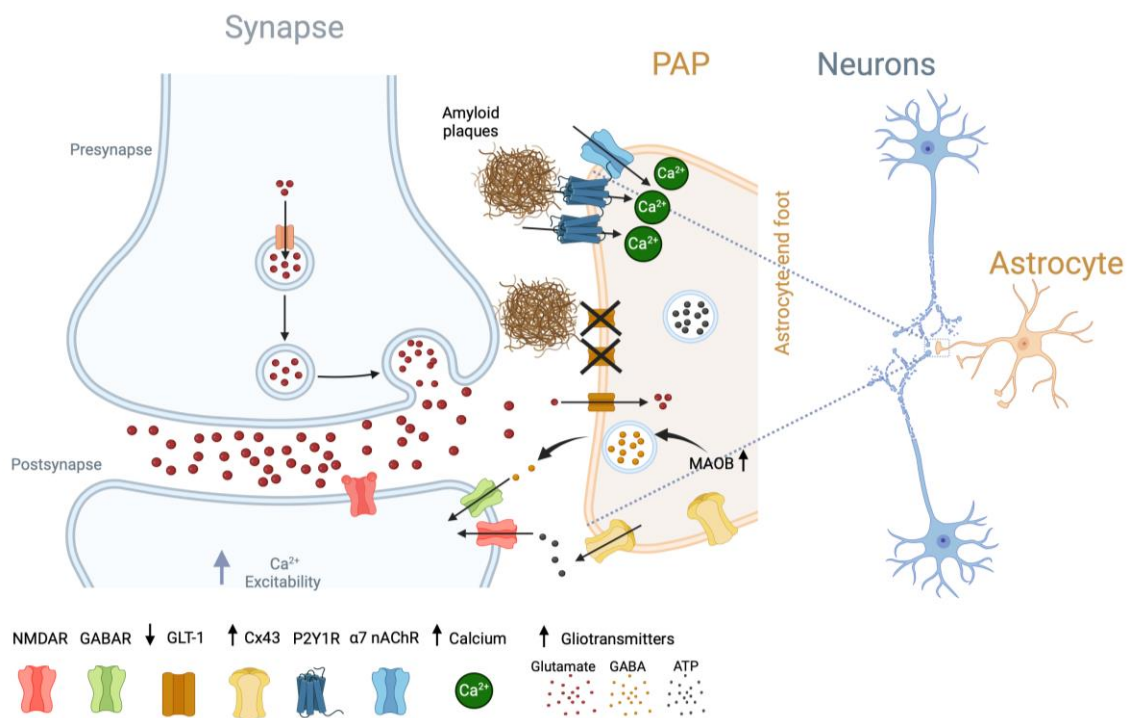


Figure 4-2 Dysregulated tripartite synapse in Alzheimer's disease
 Figure generated via [Biorender](#).

4.1.3 Aim and objectives

This chapter investigates the progressive changes in the synaptic proteome and characterises their relationship with reactive astrogliosis in AD. For this, the progressive spatial nature of phospho-tau pathology in AD was used as a proxy measure of time course, where medial regions are affected early (i.e. MTG), followed by frontal regions (i.e. PFC) and lastly, the visual cortex (VC)²⁴⁴. These spatiotemporal dynamics imply that early affected regions like MTG spend more time in a ‘disease state’, defined by the accumulation of more pathology, and thereby reflect more end-stage processes. Conversely, late-affected regions like VC may reflect more early-stage AD changes. The hypothesis is that reactive astrocytes contribute to AD pathogenesis by impacting synapse dysfunction via dysregulation in peri-synaptic processes.

To address this hypothesis, the following objectives have been set:

1. Build a spatiotemporal profile of the synaptic proteome changes in AD.
2. Characterise peri-synaptic astrocyte changes related to synaptopathy in AD.
3. Compare AD peri-synaptic astrocyte signature to signatures of reactive astrocytes defined in this thesis and the literature.

4.2 Methods

Sections 4.2.1 to 4.2.3 were not performed and written by me, but are included because they are essential for understanding the data generated and used in this chapter.

4.2.1 Synaptic and cytosolic protein extraction

This sub-section of the methods was performed and written by Jess Griffiths, a postdoctoral researcher in Seth Grant's lab.

Samples were obtained from 13 individuals with AD and 13 controls, across three brain regions: MTG, PFC and VC. Synapse-enriched and cytosolic fractions were isolated from homogenised human cortical tissue using the SynPER (Synaptic Protein Extraction Reagent) preparation protocol as per the manufacturer's instructions.

Approximately 100mg frozen tissue samples were thawed and homogenised using 15 strokes of a dounce homogeniser in Synaptic Protein Extraction Reagent (SynPER, Thermofisher) and Complete Protease Inhibitor (EDTA Free, Roche). Following homogenisation, samples were centrifuged at 1,200g for 10 minutes at 4°C. The pellet was discarded, and supernatant was transferred to a separate sample tube for centrifugation at 15,000g for 20 minutes at 4°C. The supernatant (cytosolic fraction) was extracted, snap frozen, and stored at -80°C. The remaining pellet (synapse-enriched fraction) was resuspended in cold PBS before a final centrifugation step at 15,000 x g for 20 minutes at 4°C. The supernatant was discarded, and the remaining pellet (synapse-enriched fraction) was snap-frozen and stored at -80°C.

The synapse-enriched and cytosolic fractions primarily reflect proteins derived from neuronal compartments, as the SynPER protocol is optimised for the extraction of pre- and post-synaptic elements from homogenised cortical tissue. While neurons are the dominant source of synaptic proteins, particularly those involved in vesicle trafficking, neurotransmission, and synaptic scaffolding, the cytosolic fraction may also contain contributions from other cell types present in the brain tissue—such as astrocytes, microglia, and oligodendrocytes—reflecting soluble cytoplasmic proteins released during homogenisation.

4.2.2 Label-free quantitative proteomics

This sub-section of the methods was performed and written by Harry Whitwell, the Proteomics Platform Lead at the National Phenome Centre at Imperial College London.

Synaptic proteins were eluted from synaptic vesicles by the addition of 100µL 8M urea, with 10 minutes in a sonicating water bath at room temperature, followed by centrifugation for 10 minutes at 4°C at 10,000xg. Cytosolic samples were prepared in Syn-PER buffer. A protein assay was performed on all synaptic and cytosolic fractions according to the manufacturer's instructions (DC Protein Assay microplate protocol, Bio-Rad). 20µg of protein were then taken from each sample and digested using an automated method on an Andrews+ pipetting robot (Andrews Alliance).

Briefly, samples were transferred to a digestion plate and held at 4°C while being combined with ddH₂O. A digestion buffer was then added such that the final reaction conditions were 10mM tris(2-carboxyethyl)phosphine, 40mM chloroacetamide, 100mM ammonium bicarbonate, and 0.2µg trypsin in 120µL, ensuring the urea concentration was <2M across all synaptic samples. Samples were then heated to 37°C for 16 hours before the reaction was halted by reducing the temperature to 4°C and adding 0.5% (V/V) trifluoroacetic acid (TFA).

Samples were desalted using a 96-well Oasis µElution HLB plate (Waters), where the plate was first equilibrated with 3x100µL 0.1% TFA drawn through under negative pressure. Samples were loaded onto the plate and washed with 3x100µL 0.1% (V/V) TFA before being eluted with 150µL 60% (V/V) acetonitrile and dried completely in a SpeedVac (Eppendorf) at 60°C. Peptides were dissolved in 50µL 0.1% (V/V) formic acid (FA), and a peptide assay was performed as previously described. Concentrations of peptides were then adjusted to 0.06µg/µL for analysis by LC-MS/MS.

Peptides were analysed by data-independent acquisition LC-MS/MS on an M-Class UPLC (Waters) coupled to a ZenoTOF 7600 (Sciex) in ZenoSWATH mode. 300ng were loaded onto a Kinetex 2.6µm XB-C18 150x0.3mm column (Phenomenex) for 1 minute and eluted over a 20-minute gradient, with the organic mobile phase increasing from 3% (acetonitrile with 0.1% (V/V) FA) to 30% at a flow rate of 10µL/min. After a column clean at 80% organic phase at 10µL/min for 2 minutes, the buffers returned to starting conditions of 97% aqueous phase (0.1% (V/V) FA) for 6 minutes. The column was heated to 50°C throughout.

The MS was run in positive mode, acquiring MS data between 400 and 1500Da with a 0.1s accumulation time. MS/MS data were acquired between 140 and 1800Da with dynamic collision energy enabled and an accumulation time of 0.013s. 85 SWATH windows were variably distributed between 399.5 and 903.5Da. The LC and MS were operated using OS software version 3.1 (Sciex). A study reference of pooled synaptic or cytosolic samples was run every eight samples, and a blank was run between each injection.

Proteomics raw data was searched against a predicted library of human canonical proteins using DIANN (V1.8) and annotated with sequence information from Swissprot (Uniprot). Match-between-runs was enabled.

4.2.3 Synaptome mapping

This sub-section of the methods was performed and written by Jess Griffiths, a postdoctoral researcher in Seth Grant's lab.

Synaptome mapping started with IHC on FFPE 8µm sections, including dewaxing, antigen retrieval, and blocking. Sections were incubated overnight with anti-GluN1/NR1 primary antibody, labelled with a Cy5-conjugated secondary antibody, counterstained with DAPI, and mounted for imaging. Image acquisition was conducted using imaging of GluN1⁺CY5⁺ synaptic puncta via a spinning-disk confocal microscope, followed by tile acquisition and stitching to generate large-area overview montages images. Synaptic puncta detection and segmentation were performed using an in-house Mixture of Expert ResUNet model trained on manually annotated datasets from control and late AD human cases. Detected puncta were analysed for key parameters (mean pixel intensity, size, skewness, kurtosis, circularity and aspect ratio) and classified into four subtypes using an unsupervised machine learning model. The average subtype density was quantified and averaged across all cortical layers in a 6x6mm window, excluding delineated white matter regions.

The remaining methods described in this chapter were designed and carried out by me.

4.2.4 Processing of proteomics data

The bulk proteomics synaptic and cytosolic fractions were processed using the Omix R package (v1.0)¹²⁰ to generate a comprehensive multi-assay dataset. The raw synaptic fraction comprised 6139 proteins, while the cytosolic fraction 3062. Briefly, proteomics processing was achieved via the *process_protein()* function from the Omix R package. Proteins had to be non-missing in at least 50% of samples in both the case and control groups independently to be kept; those missing in more than 50% of samples in either group were filtered out. After filtering, 5755 and 2528 proteins were kept in the synaptic and cytosolic fractions, respectively. Proteins were imputed following the minimum value method (*imputation='minimum_value'*). A batch correction was applied using median centring (*correction_method="Median_centering"*). I further filtered synaptic proteins with an inclusion list of 5667 previously identified synaptic proteins²⁴⁵ (**Supplementary Table 4-2**), yielding 3333 high-confidence synaptic proteins common to all three regions for downstream analysis.

4.2.5 Differential protein abundance

Differential protein abundance between AD and control groups was quantitatively assessed using the limma package (v.3.50.3)¹²⁴. I used the *lmFit()* function with its robust regression option to construct linear models, accommodating biological variability and potential outliers. In the regional AD versus control differential expression analysis, the model included age, sex, and PMI as covariates. Contrasts were defined and applied using the *contrasts.fit()* function, and empirical Bayes moderation was performed with robust variance estimation (*eBayes(vfit, robust = TRUE)*). P-values from the moderated t-tests were adjusted for multiple testing using the Benjamini-Hochberg procedure to control the false discovery rate (FDR)¹⁷¹. I applied a significance threshold of p-value <0.05 and Log₂FC threshold >0.2 for downstream analysis. Sensitivity analyses included continuous regressions against neuropathological measurements of amyloid and pTau to highlight similarities with the case/control setup (**Supplementary Table 4-3:4**). This was performed only on the MTG and PFC regions, as the VC regions lacked neuropathological measurements.

4.2.6 Functional enrichment

Functional enrichment analysis was performed using Enrichr¹³³. This streamlined pathway analysis allowed for the generation of pathway reports utilising common pathway databases such as KEGG, Reactome, BioCarta, MSigDB, and Gene Ontology.

4.2.7 SynGO enrichment

Regional differentially expressed proteins were analysed on the SynGo portal²⁴⁶, with brain-expressed proteins as background. SynGo enrichment plots were customised via <http://syngoportal.org/plotter#>. The Gene Ratio was derived from the ratio of GSEA counts over the GSEA pathway count included in the background (SynGo default is defined as “brain expressed”).

4.2.8 Synaptic and cytosolic fractions integration

Integration of the synaptic and cytosolic fractions was performed using the MOFA model from the MOFA2 R package (v1.4)¹⁰⁶. Before integration, scaling of the different fractions was carried out to ensure comparability. Datasets used for integration comprised the synaptic proteomics (4925 proteins), the cytosolic proteomics (2204 proteins), of which 2075 features overlapped between the two fractions. To ensure precise protein identification, ambiguous entries representing multiple isoforms or homologous proteins (e.g., “C4A;C4B”) were removed from the dataset. Unique protein identifications, such as C4B and C4A, were retained for integration. The MOFA model was configured to extract ten latent factors, comprehensively capturing variation underlying biological processes. Default settings were used for other parameters.

Factor enrichment was performed using the *run_enrichment()* function with default parameters and SynGo ontologies⁴³ as feature.sets. Factor 3 was chosen for downstream analysis, given it was enriched for synaptic gene ontologies and significantly correlated to AD neuropathology measures.

4.2.9 Pseudotime derivation

To compensate for the lack of longitudinal tissue data in *post-mortem* samples, I employed a pseudotime trajectory analysis to model synaptic-cytosolic dynamics across a disease continuum. It uses the same concept as in pseudotime inference with scRNA-seq¹⁶⁹, except here, each data point is a bulk sample derived from a patient, and pseudotime represents an inferred trajectory along which patient samples are ordered, modelling disease progression from early to advanced stages. The disease trajectory was derived using the slingshot algorithm¹³⁵ on a two-dimensional UMAP representation of AD-relevant and brain region-specific MOFA factors (Factors 1,2,3). I clustered the latent space (UMAP1/UMAP2) into 3 groups via k-means clustering and chose the root of the slingshot trajectory as the cluster with a high density of VC samples (cluster 3) and ending cluster enriched in MTG samples (cluster 2) (**Supplementary figure 4-3**). This was specified in the slingshot parameters (start.clus=3,end.clus=2). I then ranked samples according to their position along the slingshot trajectory but within predefined Braak tiers (early [Braak 0-II] and late [Braak V-VI]), implementing a semi-supervised approach. The final pseudotime values were determined by fitting a linear model between the derived disease trajectory and the sample ranks within each Braak tier. This method allows us to model AD progression as a continuous process, positioning each patient along an inferred gradient of disease severity.

4.2.10 Synaptic-cytosolic proteomics co-expression modules

Factor 3 was selected for co-expression module analysis. Factor 3 features with weights exceeding ± 1.5 SD from the mean were used to define the signature. Since this factor is negatively correlated with clinical covariates, features with negative weights increase with disease progression (“UP”-labelled features). Conversely, features with positive weights down-regulate with disease progression (“DOWN”-labelled features). Pairwise correlations between features in the UP/DOWN signatures were assessed, and a synaptic-cytosolic co-expression network was constructed where feature pairs showed an absolute Pearson’s correlation exceeding a user-defined threshold (default=0.3). This network was then dissected into functionally relevant modules using the community detection algorithm Louvain, resulting in disease-correlated modules that aid in understanding disease mechanisms. Module

eigenvalues were calculated via PCA on scaled expression data, and the first principal component was assigned as the module eigenvalue.

4.2.11 Targeted exclusion cell type enrichment

Cell-type enrichment analyses were conducted using Expression-Weighted Cell Type Enrichment (EWCE)¹³⁴ (v1.2). This analysis utilised a cell-type data reference from the Allen Human Brain Atlas¹³⁸ to determine the enrichment of specific cell types within the identified modules. To refine the characterisation of astrocyte involvement at the synaptic level, I repeated the cell type enrichment analysis after excluding genes known to be involved in peri-astrocyte processes (PAP)²⁴⁷ and neuron-astrocyte junctions (ASC)²⁴⁸.

4.2.12 Gene set enrichment

To assess the statistical overlap between these results and ASC/PAP gene sets or reactive astrocyte gene sets (*Table 4-1, Supplementary Table 4-1 & 4-10*), I used Fisher's exact tests via the *newGOM()* function of Gene Overlap package (v4.4). This test calculates a p-value to evaluate whether the observed overlap is likely if the gene sets are independent. It also provides an odds ratio, a measure of the association strength between the gene sets. All human protein-coding genes were the background set (n= 20,162).

<i>Reactive astrocyte gene sets from the literature</i>	<i>Info (i.e. check original publication)</i>	<i>REF</i>
<i>Habib_2020_DAA (n=20)</i>	Top 20 genes from Supplementary Table 2, cluster 1 vs. 4, ordered by Log ₂ FC	37
<i>Habib_2020_GFAP (n=20)</i>	Top 20 genes from Supplementary Table 2, cluster 1 vs. 6, ordered by Log ₂ FC	37
<i>Grubman_2019_A1 (n=20)</i>	Top 20 genes ordered by Log ₂ FC	225
<i>Grubman_2019_A2 (n=20)</i>	Top 20 genes ordered by Log ₂ FC	225
<i>C1_sadick (n=31)</i>	Cluster 1 from figure 5.1	222
<i>C3_sadick (n=23)</i>	C3 ⁺ cluster 5 from figure 5.B	222
<i>Mathys_2019 (n=20)</i>	Ast1 - Top 20 genes Supplementary table 6	70

<i>Reactive_consensus (n=10)</i>	Figure 1 (human)	35
<i>labib_2022 (n=7)</i>	Figure 4c	223
<i>Leng_2022_IRAS1 (n=20)</i>	Top 20 from supplementary table 4a	226
<i>Leng_2022_IRAS2 (n=20)</i>	Top 20 from supplementary table 4a	226
<i>Cameron_protective (n=20)</i>	List sent by authors - overall top 20 by Log ₂ FC by neuroprotective clusters	162
<i>Cameron_toxic (n=20)</i>	List sent by authors - overall top 20 by Log ₂ FC by neurotoxic clusters	162
<i>Clayton_Cluster2 (n=20)</i>	Supplementary table 2 - top 20 by Log ₂ FC	224
<i>Clayton_Cluster6 (n=20)</i>	Supplementary table 2 - top 20 by Log ₂ FC	224
<i>Clayton_Cluster7 (n=20)</i>	Supplementary table 2 - top 20 by Log ₂ FC	224
<i>Liddelow_A1_neurotoxic (n=12)</i>	Figure 1a	32
<i>Liddelow_A2_neuroprotective (n=13)</i>	Figure 1a	32
<i>Pan_activated (n=13)</i>	Figure 1a	32
<i>Serrano-Pozzo_2024_R1_R2 (n=20)</i>	Supplementary Data 6 R1/ R2 top 20 by Log ₂ FC	40
<i>Serrano-Pozzo_2024_AstExhausted (n=20)</i>	Supplementary Data 6 AstMet and AstTinf top 20 by Log ₂ FC	40

Table 4-1 Literature derived astrocyte gene sets

4.2.13 ASC/PAP activity

To determine ASC/PAP gene set activity in the synaptic proteomics data, I used the GSVA package (v1.42)²⁴⁹. GSVA is an unsupervised method that quantifies changes in pathway activity on a per-sample basis. I applied this analysis using the default parameters.

4.2.14 Cell-cell communication analysis

Cell-cell communication analysis was performed using NicheNet (v1.1.1)²⁵⁰. I aimed to identify and prioritise ligand proteins in astrocytes at the synapse that could potentially influence synaptic proteomics changes. NicheNet constructs a model from existing knowledge of ligand-target relationships, aiming to identify ligands produced by "sender" cells that could

influence a specific set of genes in "receiver" cells (i.e. immediate receptors and any relevant downstream gene targets). For this analysis, I used NicheNet's default settings. The "receiver" proteins included proteins from the GluN1⁺ module and its interactome (see below methods). The "sender" proteins comprised ASC and PAP proteins and the combined synaptic fractions of the ME2 glia/astrocyte module.

4.2.15 Defining the GluN1⁺ interactome

I defined the GluN1⁺ (*GRIN1* gene) interactome (i.e. protein-protein interaction network) using the STRING database (version 12.0), explicitly sourcing data from the "9606.protein.links.detailed.v12.0.txt" file. I focused on hub proteins in the synaptic fraction of the GluN1⁺ module and included only interactions with a minimum experimental score of 700, indicating solid experimental evidence. Interactions with mitochondrial proteins marked by the "MT-" prefix were excluded to prevent confounding effects.

4.2.16 Synaptic proteomics genetic enrichment

I used MAGMA software (version 1.10)¹⁷⁴ for genetic enrichment analysis of upregulated proteins. I extracted SNP locations and p-values from harmonised GWAS summary statistics^{143,175,176}. Using these SNP locations and an NCBI gene location file (NCBI38.gene.loc/ NCBI37.gene.loc, according to the GWAS), I annotated SNPs to genes using MAGMA's --annotate command. I then analysed gene-based association with reference genotypes from the 1000 Genomes Project (European population). I calculated gene-based p-values using the gene annotation file and the GWAS summary statistics. This involved calculating gene-based association results separately for three datasets (Bellenguez_magma.genes.out, Janssen_magma.genes.out, and Wightman_magma.genes.out) and then combining these results with the --meta command to generate meta-analyses p-values (meta_analysis.genes.out). The enrichment analysis incorporated these individual and combined statistics, assessing the genetic associations of different upregulated genes using MAGMA's --set-annot command.

4.2.17 Data availability

This study's metadata and patient IDs are in the Synapse repository: syn63090719. Raw mass spectrometry data are available via ProteomeXchange with the identifier PXD056052.

4.2.18 Code availability

The code for this chapter can be found in the Chapter 4 folder of the Github repository <https://github.com/eleonore-schneeg/PhD>.

4.3 Results

4.3.1 Selective vulnerability of synaptic components at different spatiotemporal stages of AD

Twenty-six *post mortem* brains were assessed using synaptomic approaches, 13 of which had been neuropathologically diagnosed with AD (Braak stage V-VI), and 13 were non-diseased control (Braak stage 0-II) cases (**Supplementary Table 4-5**). Three cortical brain regions were selected to represent regions of varying degrees of AD pathology and build a spatiotemporal profile of the synaptic proteome changes in AD (**Figure 4-3**). These regions were ordered according to the stereotypical spatial progression of AD^{24,40}, from least to highest time spent in disease state: the visual cortex (VC, Brodmann area 17), the prefrontal cortex (PFC, Brodmann area 9), and the middle temporal gyrus (MTG, Brodmann areas 20/21). The latter is affected early in AD and thereby accumulates more disease-related changes²⁴. Due to tissue availability, a subset of cases and regions were used for each experimental procedure (**Supplementary Table 4-5**). Jessica Griffiths confirmed the enrichment of synaptic proteins in the synaptic fraction using a consensus list of proteins related to individual subcellular compartments, as previously described²⁴⁵. This analysis revealed a significant enrichment of proteins localised to the post-synapse (3.3-fold, p-adjusted = 3.12×10^{-225}) and pre-synapse (1.6-fold, p-adjusted = 2.64×10^{-05}) (**Supplementary figure 4-1 a-b**). Out of a total of 5755 synaptic proteins, 77% (or 4449 proteins) were found to overlap with those reported in two recent synapse proteomic studies^{245,251} (in at least one of the two) (**Supplementary figure 4-1 c**).

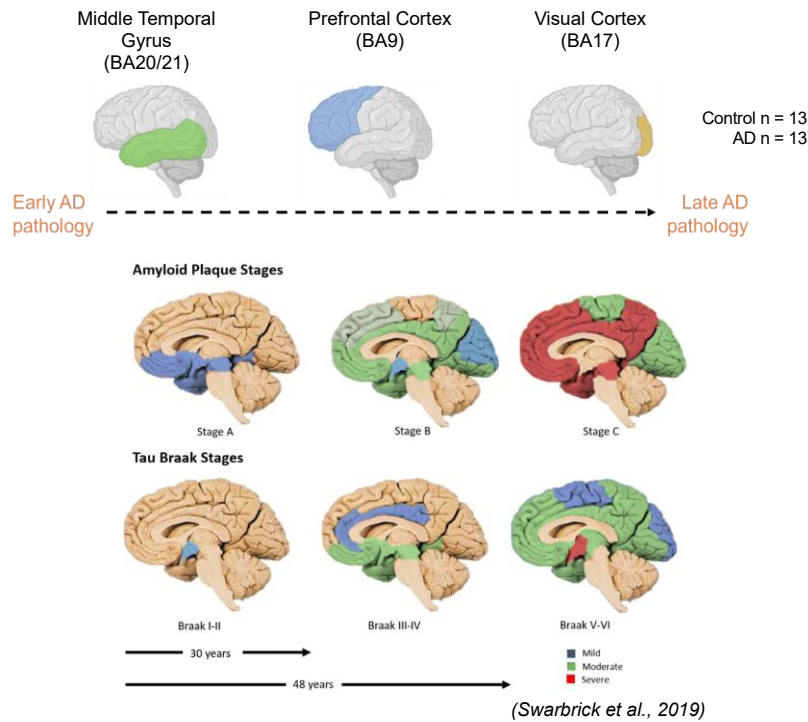


Figure 4-3 Spatiotemporal model of Alzheimer's disease

Figure adapted from Swarbirck et al.²⁵²

I next asked how many synaptic proteins were differentially expressed (DEPs) between AD and controls in the three brain regions. Differential expression analysis of 3333 proteins (those common to all three regions) revealed differences in the VC (412 DEPs: 6.9% increased, 5.4% decreased), PFC (380 DEPs: 6.6% decreased, 4.8% increased), and MTG (504 DEPs: 7.8% decreased, 7.3% increased) in AD compared to control brains (**Figure 4-4 a, Supplementary Table 4-3**). In all three regions, several canonical synaptic proteins were differentially expressed (e.g. BSN, NRXN3, SYT12), while brain region-dependent changes were also noted. For instance, GABA_a receptor subunits were downregulated in the PFC (GABRA1, GABRA2, GABARG2, GABRB3) but not in the VC or MTG. Paralog-specific changes were also observed, with membrane-trafficking synaptotagmin family members either being downregulated in all regions (SYT12), downregulated in a specific region (SYT1, SYT12 and SYT17 in the MTG), or upregulated in a specific region (SYT11 and SYT5 in the VC).

While a greater number of proteins were altered in a brain region-dependent manner, 70 proteins showed brain-region-independent alterations in AD, with 30 being downregulated and 40 upregulated. Among the proteins upregulated were astrocyte-associated proteins such as CLU, ALDH1L1, GFAP, and VIM, and complement refinement protein, C4B. Of note, the

astrocytic protein MFGE8 was upregulated in the PFC but not differentially expressed in the other regions. Downregulated proteins included postsynaptic scaffolding proteins such as BSN and PCLO, proteins involved in receptor motility like RPH3A and NRXN3, and VGF, a dysregulated protein in neurodegenerative and psychiatric disorders that has been proposed previously as a therapeutic target¹⁴⁹.

Next, I compared the abundance of proteins in subsynaptic compartments through synaptic enrichment analysis using annotations from the SynGO database²⁴⁶. DEPs were significantly represented in the synapse compartment in all cortical regions (**Figure 4-4 b, Supplementary Table 4-6**). In the VC, we noted differential enrichment between the presynaptic and postsynaptic compartments, with the most significant enrichment in the presynaptic active zone, particularly in cytoplasmic components. In the PFC, pronounced presynaptic dysfunction affected both cytoplasmic and membrane components of the active zone, along with dysfunction in the postsynaptic density, especially in its anchored component. This pathology was more widespread in the MTG, where enrichment spanned multiple pre- and post-synaptic compartments, indicating broader synaptic dysfunction.

Gene ontology (GO) pathway enrichment analysis was employed to identify biological processes associated with proteins that had a negative (**Figure 4-4 c**) or positive (**Figure 4-4 d**) fold change in AD compared to that of the control group (**Supplementary Table 4-7**). While some biological pathways were shared across regions, others represented brain region-dependent changes. For example, downregulated proteins were significantly enriched in mitochondrial-related pathways in the PFC and VC, suggesting synaptic metabolic dysfunction in these brain regions²⁵³. Additionally, we observed reduced pathways related to inhibitory synapse assembly and transmission in the PFC, although not in the VC where inhibitory synapse loss has been previously reported²⁵⁴. In the MTG, we reported a downregulation in pathways associated with synaptic organisation and function, suggesting overt synaptic dysfunction in the MTG.

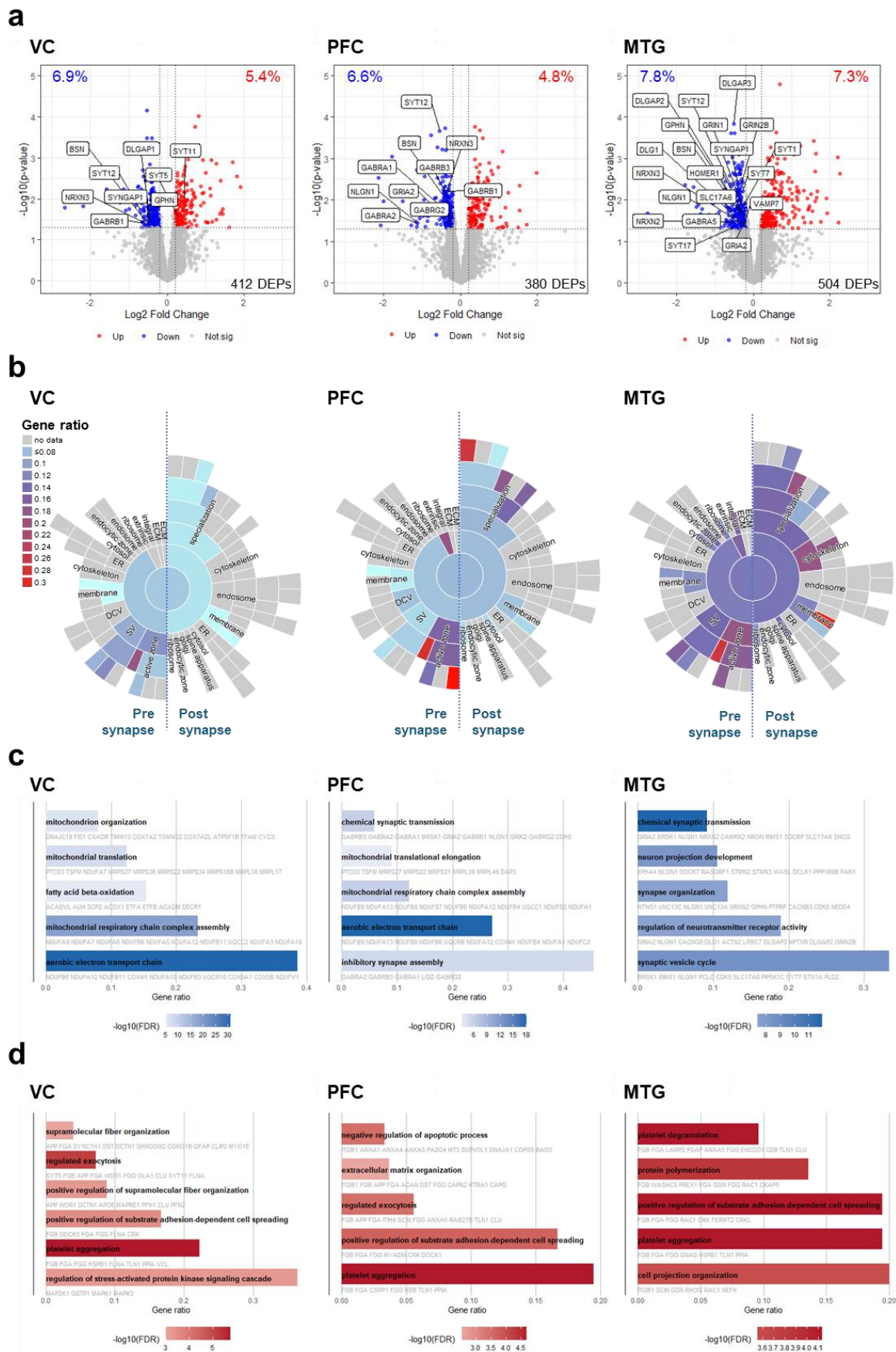


Figure 4-4 Differential synaptic protein expression across cortical regions with AD

a) Volcano plot depicting differentially expressed proteins (DEPs) between AD and control groups in the VC, PFC, and MTG. Canonical synaptic proteins of interest are highlighted. **b)** Sunburst plots displaying the gene ratio calculated as the ratio of significant genes versus all annotated genes per SynGO cellular component term in the VC, PFC, and MTG (one-sided over-representation Fisher's exact test with multiple testing correction using FDR). **c,d)** Bar plots illustrating the relative functional enrichment against GO biological process terms on DEPs that have a negative **c)** and **d)** positive fold change in the VC, PFC, and MTG in AD relative to control (colour, $\log_{10}(\text{FDR})$, one-sided over-representation Fisher's exact test).

I further examined the relationship between these differentially expressed synaptic proteins and AD risk genes from three GWAS studies^{143,175,176}. Using MAGMA, a gene set analysis tool for GWAS data, I found that upregulated DEPs were enriched in AD risk genes across all regions (**Figure 4-5 a**). Most proteins driving the enrichment were unique to regions, with the MTG showing 39 unique proteins. (**Figure 4-5 b**). Seven common proteins were found to drive AD genetic risk across the three regions, including CLU and GFAP, markers of activated astrocytes. Looking at all proteins driving genetic risk in synaptic proteomics regardless of region, I found pathways that include the regulation of amyloid fibril formation and neuron projection development, showcasing well-known AD risk genes like *APP* or *APOE* (**Figure 4-5 c**).

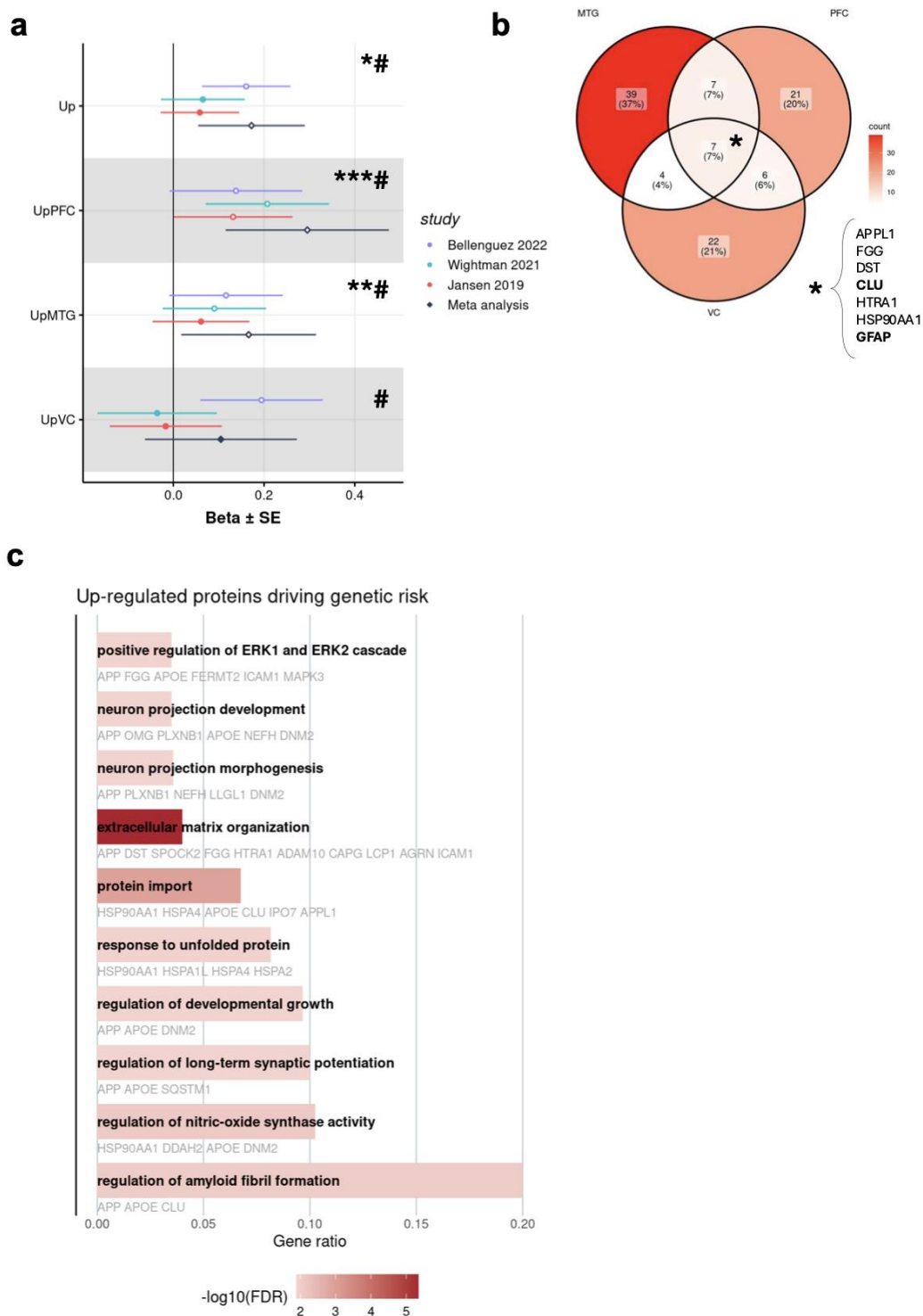


Figure 4-5 Up-regulated proteins at the synapse are enriched in AD genetic risk
a) Forest plot illustrating the enrichment of upregulated proteins (all regions: PFC, MTG, VC) in genetic risk quantified by three AD GWAS and in a meta-analysis, highlighting significant beta values and corresponding P-values. # indicates modules enriched in AD risk genes in the meta-analysis and * only in one or more studies. **b)** Venn Diagram showing the overlap of up-regulated proteins driving the GWAS enrichment (meta-analysis) in the three regions. **c)** Top pathways of synaptic proteins found to drive AD genetic risk regardless of the brain region.

4.3.2 Differential excitatory synapse subtype vulnerability in AD

This sub-section of the results was written by Jess Griffiths, a postdoctoral researcher in Seth Grant's lab, but it remains integral to the overall findings and needed for the understanding of integrative results.

To understand how changes in the synapse proteome may be associated with synapse loss and vulnerability, Jessica Griffiths next assessed synaptic protein expression using immunolabelling of tissue sections from AD and control cases (**Supplementary figure 4-2**). She focussed on the NMDA receptor subunit GluN1, the obligatory subunit of this ionotropic receptor, which plays a key role in synaptic plasticity and learning⁵¹. GluN1⁺ synaptic puncta were immunolabelled, imaged, quantified and classified into four morphological subtypes (S1-S4) based on their punctum intensity and shape parameters (**Figure 4-6 a-b**), as similarly conducted in the mouse brain^{52,53}. The total GluN1⁺ puncta density (number per 100 μm^2) was quantified across the cortical grey matter in each brain region (**Figure 4-6 c**). An unpaired t-test revealed no significant change in GluN1⁺ synapse puncta density in the VC (p-value = 0.42); however, there was a significant decrease in the PFC (p-value = 0.03) and MTG (p-value = 0.02) in AD compared to control. Next, she examined the GluN1⁺ subtypes and found a significant loss of S2 and S4 excitatory synapse subtypes, an increase in the S3 subtypes, and no change in S1 in MTG and no changes in the PFC or VC. This suggests that although not all NMDA receptor-expressing synapses in the MTG are lost, certain specific excitatory synapse subtypes are particularly susceptible to loss.

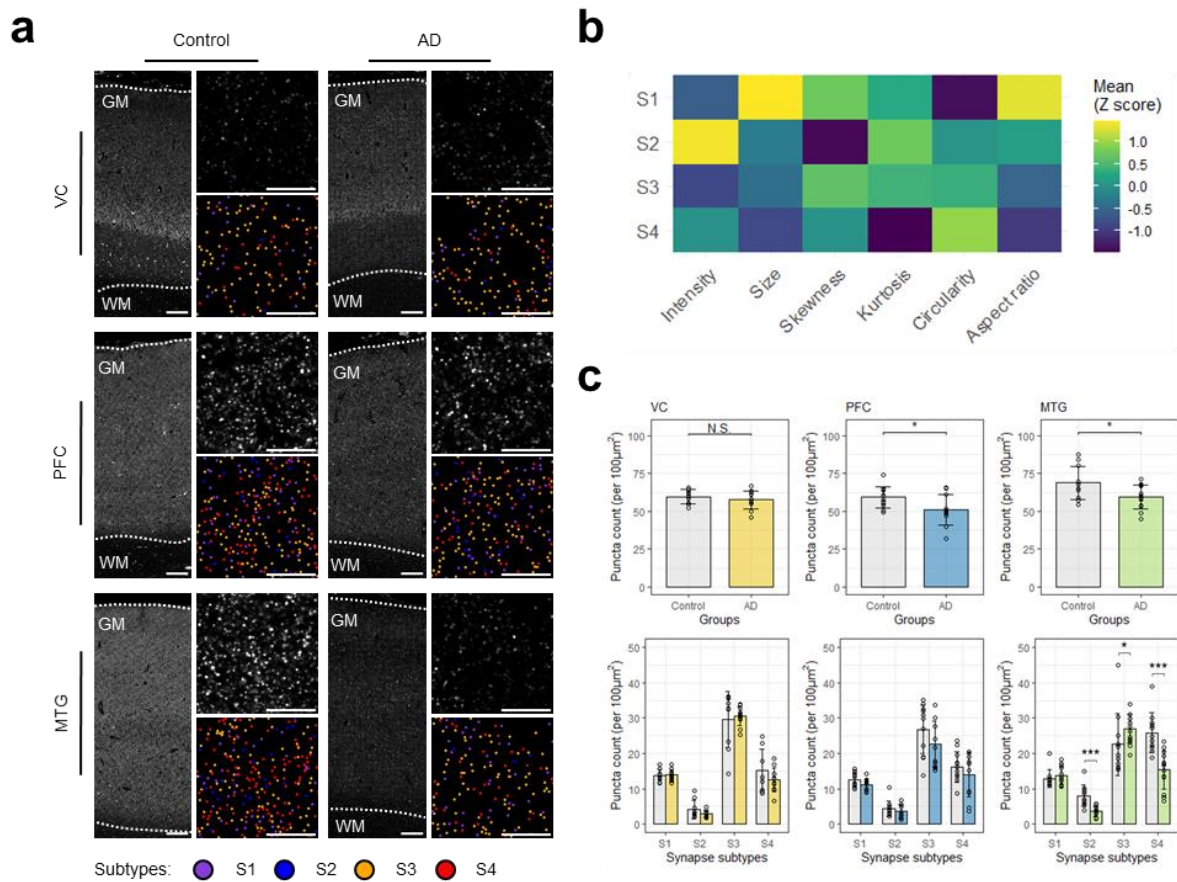


Figure 4-6 GluN1⁺ synapse subtypes are differentially affected by AD pathology in the MTG

a) Representative immunohistochemical images showing GluN1⁺CY5⁺ synaptic puncta in three cortical brain regions in AD and control cases. High-magnification images taken from superficial layer II of the cortex with mapped GluN1⁺ synapse subtypes illustrated below (S1-S4) (GM, grey matter; WM, white matter; scale bar, 10µm). **b)** Mean values of synaptome parameters (puncta intensity, size, skewness, kurtosis, circularity, aspect ratio) for GluN1⁺ subtypes. Values are normalised as Z-scores for each parameter. **c)** Bar plots displaying mean densities of GluN1⁺ synaptic puncta (unpaired *t*-test, **p*≤0.05) and densities of GluN1⁺ synapse subtypes (multiple unpaired *t*-tests or Mann-Whitney *U* tests with Benjamini-Hochberg correction, **p*≤0.05, ***p*≤0.01, ****p*≤0.001) in AD and control groups. Jittered dots represent individual case values.

4.3.3 Synaptic-cytosolic proteomics integration identifies a divergent glia-synapse relationship

To further understand the mechanisms that co-ordinate cell-wide (synaptic and cytosolic) and local synaptic changes in relation to disease progression and synapse subtype pathology, I conducted an unsupervised integration of the synaptic and cytosolic proteomics using the MOFA model¹⁰⁶ to identify shared axes of variation (i.e. *factors*) relevant to AD pathology.

Using K=10 factors, the model explains up to ~23% of the variation in the synaptic fraction and ~41% in the cytosolic fraction (**Figure 4-7 a, Supplementary Table 4-8**). I defined the AD-correlated factors as 1,2,3, given the significant correlation between AD neuropathology variables and these factors (range of R^2 value from 0.38 to 0.64, $p \leq 0.05$, **Fig 5b**). These AD-correlated factors were used in a semi-supervised approach to derive a pseudotime measurement, representing disease progression, where a higher pseudotime represents a longer time spent in the disease state. Briefly, this involved deriving a disease trajectory from the UMAP of AD-correlated factors (1-3) and then ranking samples according to this trajectory within early (0-II) and late (V-VI) Braak tiers (**Supplementary figure 4-3**). Concerning the vulnerable GluN1⁺ synapse subtypes in AD, I observed that pseudotime was significantly negatively correlated with the summed density of S2 and S4 synapse subtypes in the MTG region (**Figure 4-7 c**). In contrast, an opposite trend was found for the S3 subtypes. I next selected factor 3 for downstream analyses because its variation was derived from both synaptic and cytosolic fractions, unlike factor 1, primarily explained by the cytosolic fraction, and factor 2, mainly explained by the synaptic fraction. Specifically, factor 3 accounted for approximately 5% of the variance in the synaptic fraction and 3% in the cytosolic fraction. More so, factor 3 showed significant enrichment in SynGO pathways, indicative of synaptic function changes related to AD (**Supplementary figure 4-4**). Constructing synaptic-cytosolic co-expression networks based on top features driving factor 3 (weights ± 1.5 SD from the mean) revealed three AD-correlated modules (**Figure 4-7 d**). Two of these modules were prioritised for further analysis due to their strong associations with AD pathology and lack of age confound (**Supplementary Table 4-9**). Using EWCE to test for cell-type enrichment against a reference dataset from the Allen Human Brain Atlas, I found that F3_DOWN_ME4 was related to neuronal cell types, whereas F3_UP_ME2 was enriched for glial cells, including astrocytes (**Figure 4-7 d**). To ensure baseline regional differences did not confound these modules and truly reflected AD-specific changes, I assessed case-control module eigenvalue differences stratified by brain region (**Figure 4-7 e**). For the F3_UP_ME2 astrocytic module, AD vs. controls differences increased, with a regional trend as expected spatiotemporally (VC<PFC<MTG) in AD but constant in controls. Similarly, for the F3_DOWN_ME4 module, eigenvalues show a clear separation between AD and control within regions and a trend as expected spatiotemporally (VC<PFC<MTG), which indicates that the module captures disease-driven changes. In parallel, paired t-tests indicated no differences between brain

regions, further strengthening the idea that AD regional trends (i.e. not observed in controls) are attributed to AD sub-staging of the disease rather than baseline regional differences (*Supplementary figure 4-5*). The F3_UP_ME2 astrocytic module was strongly negatively correlated to the F3_DOWN_ME4 synaptic module ($R=-0.87$, $p = 2.2e-16$, *Supplementary figure 4-6*).

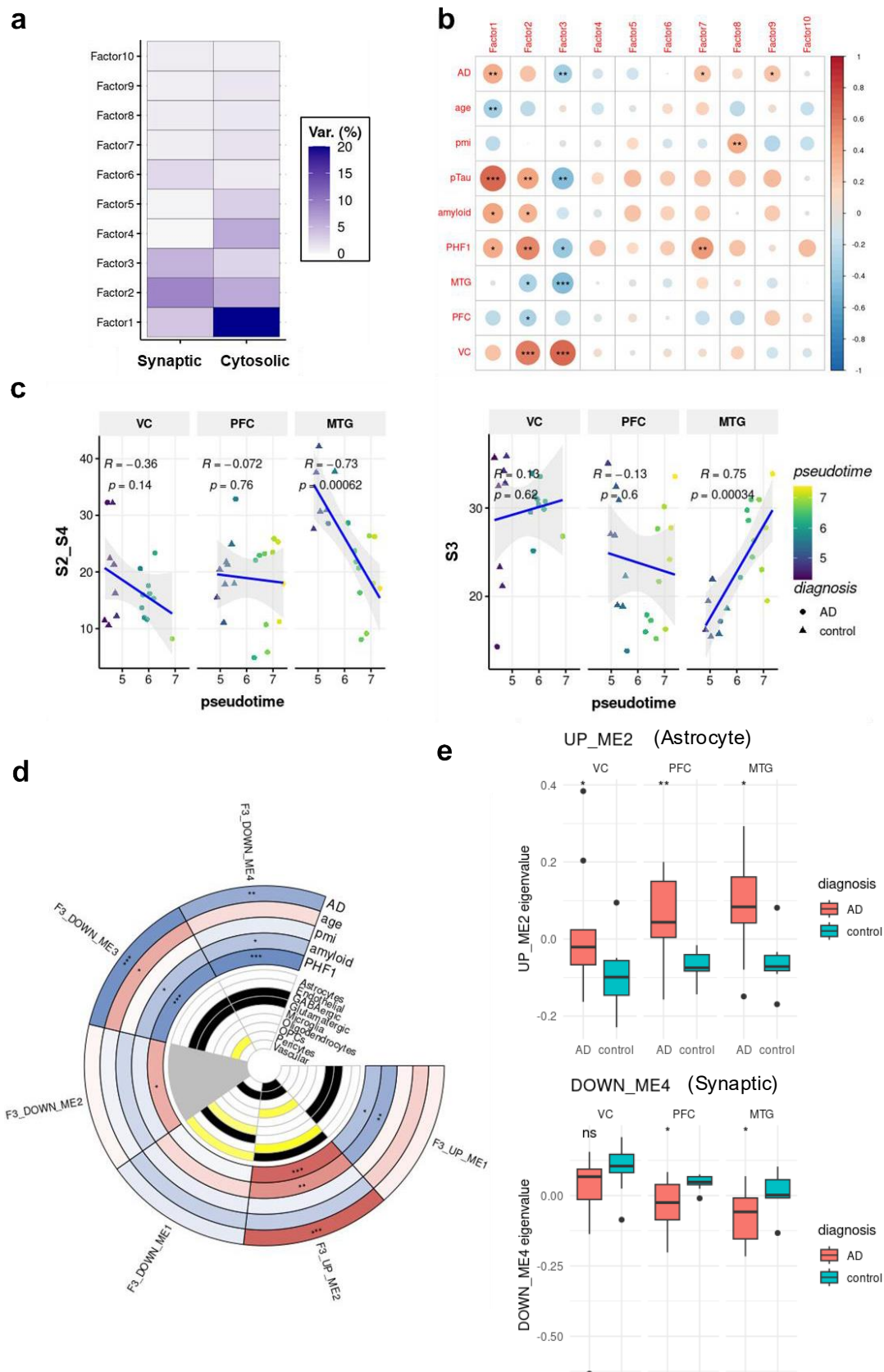


Figure 4-7 Synaptic-cytosolic proteomics integration unravels divergent glia-synapse dynamics

*a) Variance decomposition showing the percentage of explained variance per proteomics synaptic and cytosolic fractions and factor of the MOFA model with 10 factors. For each fraction, the heatmap shows the percentage of variance explained by the respective factor. b) Heatmap showcasing the Pearson's correlation matrix of clinical and pathological parameters with 10 MOFA factors derived from bulk synaptic-cytosolic integration. c) Linear regression of the summed densities (count per 100 μm^2) of S2 and S4 synapse subtypes (S2_S4) and of S3 subtype against pseudotime, stratified by brain regions. d) Circos plot visualising module (ME) trait relationships Pearson's correlations indicating significant associations with key AD features (positive association in red, negative in blue). Cell type enrichment in the inner circle indicates significant enrichment in black ($p \leq 0.05$), near significance in yellow ($p > 0.05 - p < 0.1$), and non-significant in white. e) Boxplots assessing AD vs. controls module eigenvalue differences, for the synaptic and astrocyte module, stratified by brain regions (Wilcoxon rank sum test, * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$).*

I further investigated the proteins in F3_DOWN_ME4, which primarily includes proteins from the synaptic fraction, such as GluN1 (labelled with *GRIN1*) (**Figure 4-8**), also used in the synaptome mapping (**section 4.3.2**). Linear regression analysis of the summed density of S2 and S4 synapse subtypes as a function of the module eigenvalue revealed a downregulation of the module corresponding to the decrease in the summed S2 and S4 density, specifically in the MTG region. This relationship mirrored the abundance of the GluN1 protein and summed S2 and S4 density, underscoring the critical relationship between GluN1 and synapse subtypes (**Figure 4-8 b**). While an outlier with low synapse density was noted in the VC region, this region was not significant in the analysis, and the presence of the outlier did not affect the overall interpretation of the MTG-specific association. Additionally, I observed a significant enrichment of synaptic transmission pathways within this module (**Figure 4-8 c**). This module indicates localised regulatory mechanisms or protein networks confined to the synaptic compartment, involving proteins essential for neurotransmitter release, synaptic receptor dynamics, and synaptic scaffolding, without counterparts or direct interactions with cytosolic proteins. The GluN1⁺ module highlights the presence of highly specialised synaptic processes that operate independently of broader cellular changes, emphasising the compartmentalised nature of synaptic regulation in the context of AD pathology.

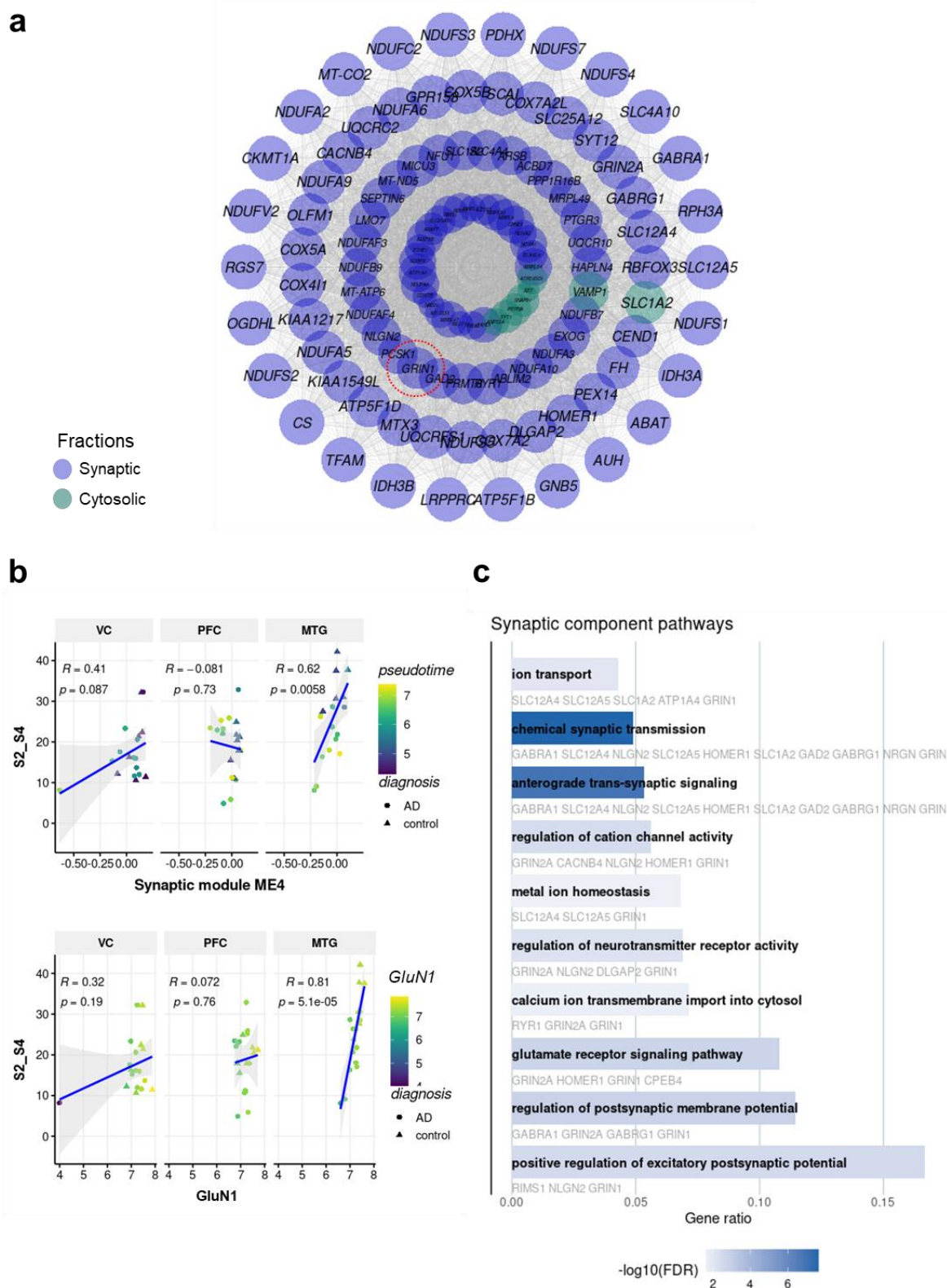


Figure 4-8 Summed S2 and S4 synapse subtype loss is associated with synaptic proteomic differences

a) Schematic co-expression network diagram showing module membership of protein from the cytosolic fraction (green) and protein from the synaptic fraction (purple) in the GluN1^+

(GRIN1; red circle) synaptic module. Node size is proportional to connectivity (node with a degree >50 shown on the graph for clarity), and nodes are separated into four quantiles representing each layer. **b)** Linear regression of summed density of S2 and S4 synapse subtypes (S2_S4) as a function of the synaptic module ME4 eigenvalue (F3_DOWN_ME4) (top) and GluN1 protein abundance (bottom). **c)** Bar plots of the functional enrichment analysis against GO biological process terms (containing GluN1 [GRIN1]) on the synaptic fraction proteins. Gene ratio represents the ratio of significant genes versus all annotated genes per pathway (colour: $\log_{10}(\text{FDR})$ one-sided over-representation Fisher's exact test).

In contrast, the exploration of the F3_UP_ME2 glial module displayed proteins from both the synaptic and cytosolic fractions (**Figure 4-9 a**). Linear regression analysis of the summed density of S2 and S4 synapse subtypes as a function of the glia module eigenvalue confirmed a significant relationship (**Figure 4-9 b**). Next, I wanted to characterise peri-synaptic astrocyte changes related to synaptopathy in AD. To determine if the astrocyte enrichment observed in the synaptic fraction could be attributed to proteomic changes associated with PAP and neuron-astrocyte junction proteins (ASC), I conducted cell-type enrichment analyses for both fractions with and without PAP/ASC proteins (**Figure 4-9 c, Supplementary Table 4-1**^{247,248}). Astrocyte enrichment in the synaptic fraction was no longer significant when PAP/ASC proteins were removed, unlike in the cytosolic fraction. This suggests the presence of other astrocytic components or functions within the cytosolic milieu, and that the astrocytic signature in the synaptic fraction is explained mainly by changes in proteins localised in peri-synaptic processes and junctions. The synaptic fraction of this module, which consistently overlapped with the ASC junction module identified by Niu et al.²⁴⁸ (**Supplementary figure 4-7**), showed enrichment in pathways related to the regulation of neuron death, receptor-mediated endocytosis, and regulation of the complement system (**Figure 4-9 d**). In contrast, the cytosolic fraction had an upregulation of pathways associated with cell motility and response to unfolded protein (**Figure 4-9 e**). NicheNet analysis was used to determine ligand-target interactions between astrocytes and neurons converging at the synapse. The ligand with the potential to regulate the most targets was the upregulated DKK3, likely acting through its receptor KREMEN1 (**Figure 4-10**), both of which have been implicated in synapse loss^{255,256}. NicheNet analysis provided a number of potential synaptic targets which DKK3 are likely to impact. Taken together, these provide a basis for understanding the intricate interactions and signalling pathways between peri-synaptic astrocytes and synapses in AD.

Figure 4-9 Astrocyte proteome enrichment at the synapse is dominated by astrocyte junction (ASC) and peri-synaptic astrocytic process (PAP) proteins

a) Schematic co-expression network diagram showing module membership of protein from the cytosolic fraction (green) and protein from the synaptic fraction (purple) in the astrocyte module. Node size is proportional to connectivity (node with a degree >50 shown on the graph for clarity), and nodes are separated into four quantiles representing each layer. **b)** Linear regression of summed density of S2 and S4 synapse subtypes (S2_S4) as a function of the glia module eigenvalue. **c)** EWCE cell type enrichment of ME2 Glia module, stratified by cytosolic and synaptic fractions. Red bars indicate enrichment results on the complete cell-type data reference from the Allen Human Brain Atlas. Grey bars indicate enrichment results on the same reference data but where PAP and ASC genes were filtered out. Enrichment p-values adjusted for multiple testing via Bonferroni correction (ns>0.5, * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$). **d,e)** Bar plots of the functional enrichment analysis against GO biological process terms on the synaptic fraction (**c**) and cytosolic fraction (**d**) proteins. Gene ratio represents the ratio of significant genes versus all annotated genes per pathway (colour: $\log_{10}(\text{FDR})$ one-sided over-representation Fisher's exact test).

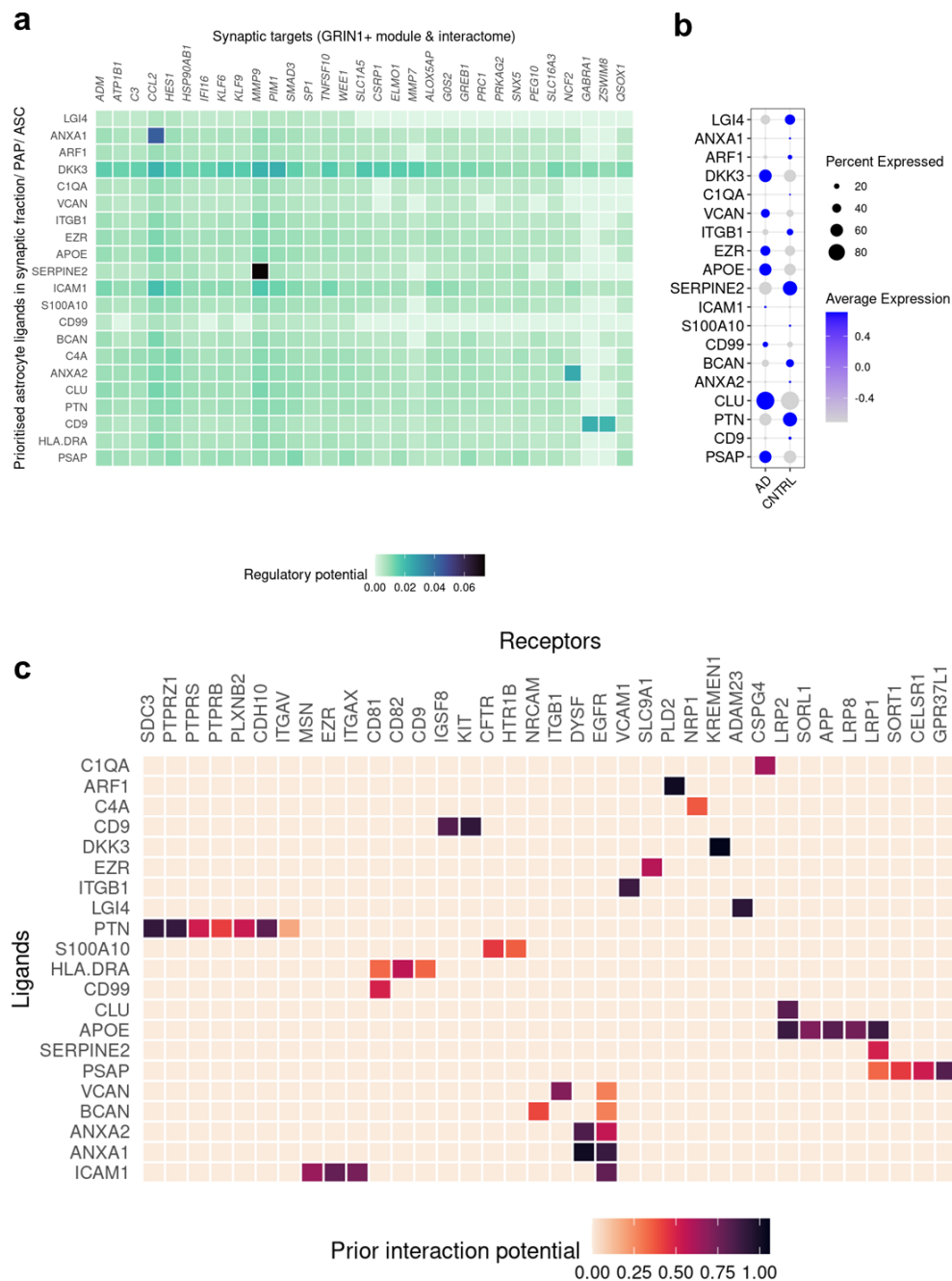


Figure 4-10 Astrocyte to synaptic targets communication

a) Heatmap showing NicheNet's ligand–target matrix denoting the regulatory potential between astrocyte module ME2 / PAP / ASC ligands and target genes in the expanded *GluN1*⁺ (*GRIN1*) module. **b)** Average expression of prioritised astrocyte ligands in astrocyte single nuclei data. Size indicates the percentage of cells expressing the genes, and colour indicates average scaled expression. **c)** Heatmap showing NicheNet's predicted upstream ligand–receptor pairs that might regulate genes in the extended *GluN1*⁺ module (module + interactome).

4.3.4 Astrocyte signature at the synapse versus general reactivity profiles in AD

I next asked whether the AD peri-synaptic astrocyte signature characterised was also related to overall astrocyte reactivity, as defined in the literature and **Chapter 3** of this thesis. I extracted reactive astrocyte signatures from Liddel³², Habib³⁷, Sadick²²², and others^{40,70,162,223–226}. (**Supplementary Table 4-10, Supplementary figure 4-8**) and compared them to the AD peri-synaptic astrocyte signature (UP_ME2_SYN) and other reactive astrocytes defined in this thesis, including (1) F14_UP_ME2 and F11_UP_ME1 from bulk transcriptomics-proteomics integration in **Chapter 3**; (2) Single nuclei cluster Astro3 (Intermediate reactive) and Astro2&5 (classical reactive) regressed against pathology measure of A β and PHF1 in **Chapter 3**; (3) The cytosolic fraction of UP_ME2, defined in this chapter, UP_ME2_CYT (**Supplementary Table 4-11**).

Overall, I found that the AD peri-synaptic astrocyte signature UP_ME2_SYN was enriched in multiple literature-defined reactive astrocyte gene sets and distinctively more in the consensus reactive signature (**Figure 4-11**). Moreover, this signature was consistent with the reactive bulk signature F11_UP_ME1 defined in **Chapter 3** and overlapped with both neurotoxic and neuroprotective signatures (**Figure 4-11**). **Figure 4-12** summarises the pathways enriched in all the literature-defined sets into broad categories, including “complement”, “TNF- α and cytokine signalling”, “interferon-gamma response”, “cell proliferation” and “apoptosis”. Pathways enriched in the peri-synaptic astrocyte signature (UP_ME2_SYN) were then filtered for terms that belong to these broad categories, as a way to highlight potential mechanisms and targets through which reactive astrocytes contribute to synaptopathy in AD (**Supplementary Table 4-12**). Top pathways included cytokine mediated signalling (FDR= 0.00125), regulation of complement activation (FDR=0.00125), calcium signalling (FDR=0.037) and regulation of neuron death (FDR=0.039).

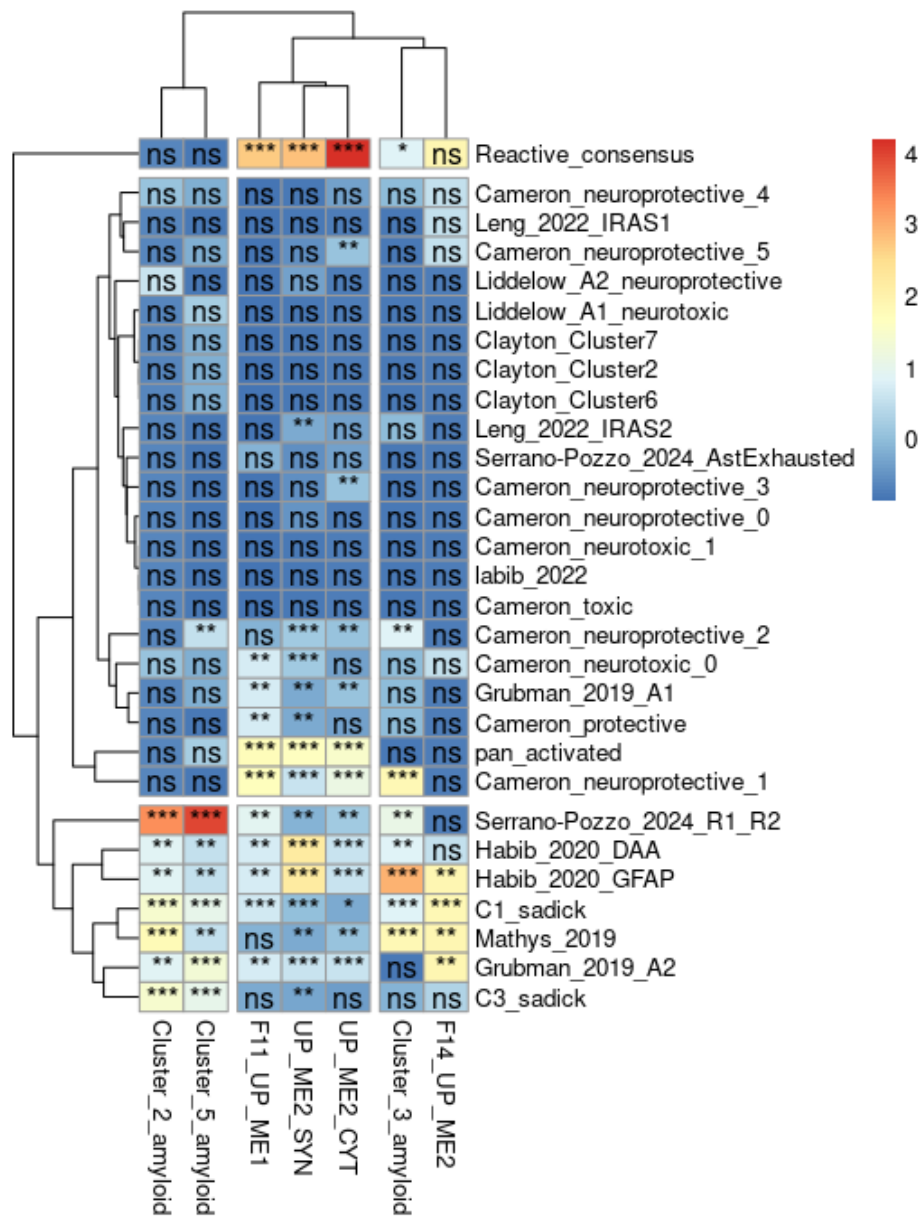


Figure 4-11 Consensus reactive astrocyte signatures derived in this thesis versus the literature

Heatmap illustrating the enrichment of thesis-defined reactive astrocyte gene sets compared to literature-defined signatures. Columns represent astrocyte signatures defined in this thesis, including peri-synaptic proteomics (UP_ME2_SYN/CYT), Chapter 3-defined bulk modules (F14_UP_ME2 and F11_UP_ME1) and genes up-regulated with increased amyloid (“_amyloid”) from Astro3 (Intermediate reactive) and Astro2&5 (classical reactive). Rows represent curated astrocyte gene sets from the literature. The heatmap is hierarchically clustered to highlight relationships across signatures. Distinct literature clusters suggest a consensus reactive astrocyte profile and delineate reactive sub-states. The colour scale indicates column-scaled odds ratios of enrichment (* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$).

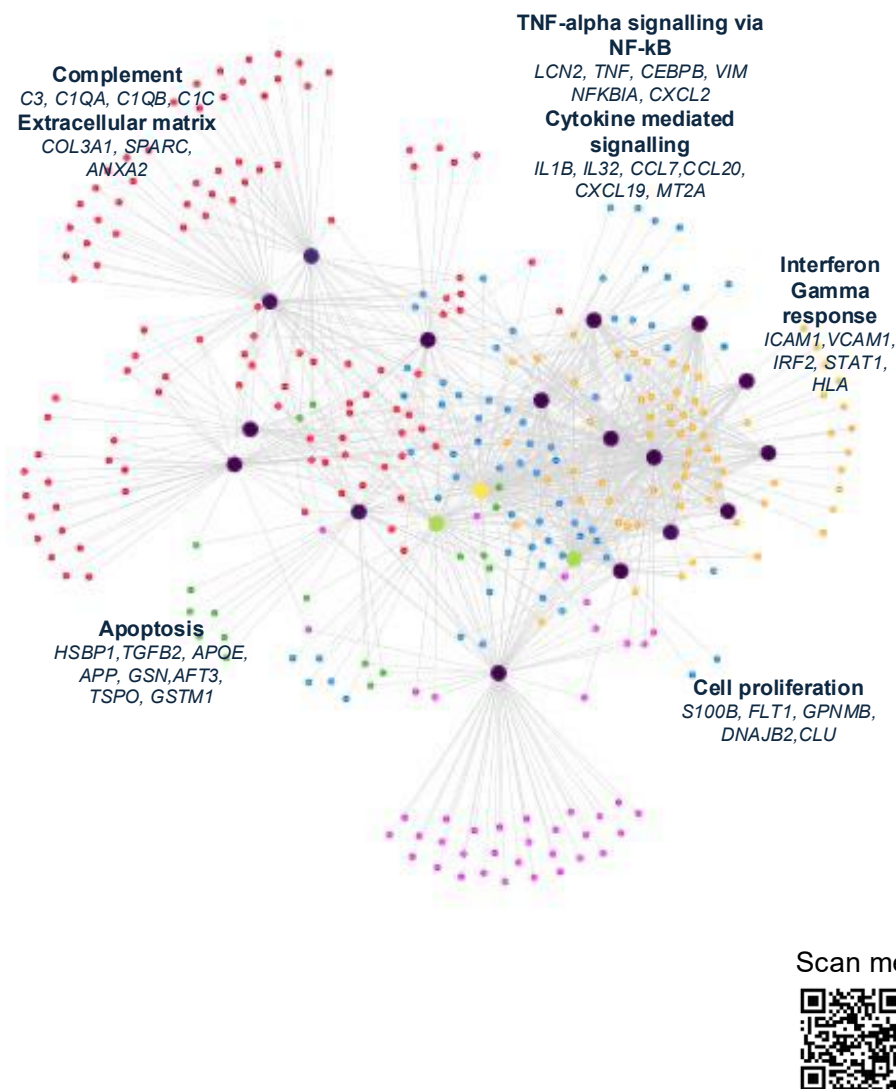


Figure 4-12 Consensus reactive astrocyte signatures interactive network of pathways and genes

The interactive network from the Omix Pathway Report is accessible via the QR code or the following [link](#). Clusters or pathways groups are coloured in red, blue, yellow, pink, and green.

4.3.5 Linking astrocyte reactivity to synaptic resilience

Having related the AD peri-synaptic astrocyte signature to overall astrocyte reactivity, I wanted to study its relationship to synaptic resilience, that is, to assess whether the extent of synapse loss could be related to the extent of astrocyte reactivity at the synapse. S2 and S4 subtypes were previously defined as those with preferential loss in AD, specifically in the MTG region as reported in section 4.3.2. I explored the relationship between S2_S4 summed density and global pathology (as the sum of $\log_{10}(\text{PHF1})$ and $\log_{10}(\text{A}\beta)$ loads) in the MTG to derive a slope of expected synapse density according to pathology levels, similarly to what was done in

section 3.3.8. A synaptic resilience score was then derived as the difference in observed minus expected synapse S2-S4 density (fitted values on the line) and related to UP_ME2 AD peri-synaptic astrocyte signature values. In this model, higher astrocyte reactivity at the synapse was associated with worse-than-expected synaptic resilience (or higher synaptic loss than predicted from global pathology). Altogether, GluN1⁺ synapse subtypes S2_S4 loss in the MTG could be worsened by astrocyte reactivity, suggesting a detrimental role at the synapse ($R = -0.54$, $p\text{-value} = 0.045$) (**Figure 4-13**). Taken together, these results provide a basis for understanding the interactions and signalling pathways between peri-synaptic astrocytes and synapses in AD.

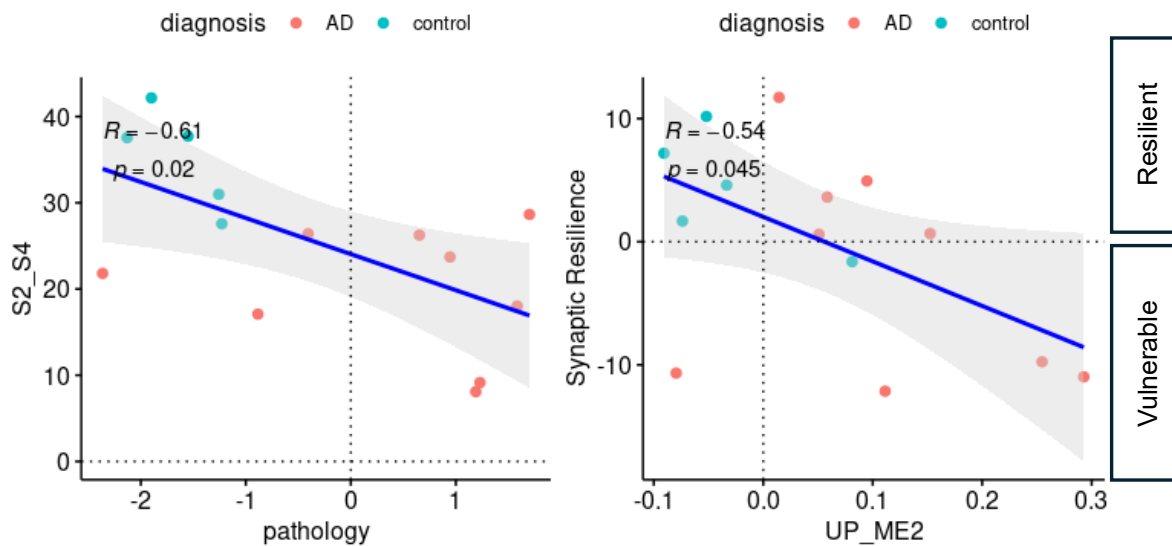


Figure 4-13 Astrocyte reactivity at the synapse worsens S2_S4 synaptic resilience
Left) Scatter plot depicting the relationship between S2_S4 synapse density and global pathology (sum of $\log_{10}(\text{PHF1})$ and $\log_{10}(\text{amyloid})$) in the MTG region, with a fitted regression line representing expected synapse density as pathology increases. This line serves as the baseline for synaptic resilience calculation, against which deviations are assessed to determine resilience scores. **Right)** Scatter plot showing the relationship between synaptic resilience (observed minus expected S2-S4 synapse density) and UP_ME2 AD peri-synaptic astrocyte signature values.

4.4 Discussion

Synapse loss is strongly correlated with cognitive impairment in AD^{53–56}. However, we have yet to determine how synaptic dysfunction and subsequent loss progress in the disease. Here, we have conducted a synaptoproteomics study using multiple brain regions from AD and NDC human *post mortem* brain tissue to build a spatiotemporal profile of synaptopathy and identify

vulnerable synaptic components. We identified potential therapeutic mechanisms by determining synapse-specific changes and elucidating the role of peri-synaptic astrocytic processes in synaptopathy.

4.4.1 Synaptic proteomics spatiotemporal changes

The loss of synapses is a well-documented hallmark of AD, yet the events preceding synaptic loss remain relatively unexplored, especially in human AD. This study aimed to characterise these early synaptic changes using synaptic proteomics and synaptome mapping across multiple brain regions in post-mortem AD tissue, representing the stereotypical spatial progression of AD^{24,40}. Within-region case-control analysis isolated AD-specific molecular alterations and allowed us to capture spatiotemporal relevant changes without confounding effects from inherent baseline differences across regions. The findings revealed that most differentially expressed proteins in AD were unique to specific brain regions, with minimal overlap, indicating spatiotemporally-stamped mechanisms as AD progresses. In the VC, a region with the lowest pathology and indicative of early-stage disease, we observed evidence of metabolic downregulation, including mitochondrial dysfunction, a well-known early-stage AD characteristic²⁵⁷. Interestingly, we also detected upregulation of exocytosis, possibly due to increased presynaptic vesicle release. This could be explained by mitochondrial dysfunction, which increases the number of vesicles in the presynaptic active zone and reduces the threshold for vesicle release²⁵⁸, thereby enhancing exocytosis. These early mitochondrial changes have also been implicated as precursors to hyperexcitability²⁵⁹, which may be an early marker of synaptic dysfunction in AD. In the PFC, a region typically affected at mid-stage AD²⁶⁰, there was a downregulation of inhibitory synapse assembly, evidenced by reduced levels of GABAergic receptor subunits. This finding aligns with the emergence of hyperexcitability in AD, first observed during this stage of the disease²⁶¹. The PFC also exhibited downregulated mitochondrial function and upregulated exocytosis pathways, suggesting continued elevated vesicle release, further contributing to hyperexcitability. In the MTG, the region most affected by AD pathology, characterised by the greatest duration in the disease state, we found broad reductions in canonical synaptic transmission pathways and decreased GluN1⁺ synapse density. This is consistent with previous findings of reduced PSD-95 levels in the temporal cortex and disruption of the excitatory-inhibitory balance and associated with cognitive impairment in

AD²⁶². Alongside distinct regional signatures, brain region-independent protein changes were also seen, with progressive differential expression across the visual, prefrontal, and temporal cortices. These included upregulated astrocytic markers and downregulated presynaptic scaffolding proteins. While the focus was on statistically significant changes, we also observed proteins showing a consistent direction of effect across regions but not reaching significance—possibly due to smaller effect sizes. We speculate that these may represent earlier-stage alterations that precede the more robust changes detected later in disease progression. Finally, the findings indicated that pre-synaptic proteins are affected earlier in the disease process, particularly in the VC region. At the same time, post-synaptic changes become more prominent as the disease progresses, especially in the PFC and MTG regions. This progression of synaptic changes in human AD contrasts preclinical models where post-synaptic components appear more resilient²⁰⁶.

4.4.2 The impact of PAPs in AD

Integrating data from synaptic and cytosolic fractions allowed us to investigate the autonomy of synapses in the context of AD. I identified a synaptic module DOWN_ME4, predominantly driven by the synaptic fraction, comprising the GluN1 protein, which decreased as the disease progressed. This decline in the synaptic module aligns with worsening synaptic health and synapse subtype loss rather than suggesting resilience in the remaining synapses. This synaptic signature was strongly inversely related to UP_ME2, subsequently found to be enriched for astrocytes and mostly driven by PAP, such as Ezrin (EZR)²⁶³, and ASC proteins in the synaptic compartment. This is consistent with a previous study highlighting the inverse correlation between changes in synaptic functions and PAP²⁶⁴. This signature included the upregulation of endocytosis-related proteins necessary for astrocytic neurotransmitter uptake²⁶⁴. The astrocytic module also included upregulation of complement activation components, such as C1Q and C4, which are implicated in astrocytic engulfment of synapses²⁶⁵. C1Q is involved in the initiation of complement cascades that lead to synapse pruning usually associated with microglia, while C4, primarily secreted by astrocytes²⁶⁶, is known to be increased in AD and is associated with synaptic engulfment by astrocytes²⁶⁷. Even though C3 was not present in the module, the synaptic fraction was enriched for the C3⁺-astrocyte cluster defined in Sadick et al., but not the cytosolic fraction. C3⁺-astrocytes have neurotoxic function and are

predominantly found in regions of neurodegeneration²²², thereby here suggesting a neurotoxic astrocyte signature localised to the synaptic fraction, while the cytosolic proteins within this module were mostly linked to processes such as cell motility²⁶⁷ and responses to unfolded protein stress, consistent with the characteristics of reactive astrocytes that lose their supportive role for synapses in AD. These results also align with the recent observation that p62/SQSTM1⁺ astrocytes, a feature in this analysis' module, are associated with synaptic loss and dysregulated calcium signals⁴⁵. These have been found to exhibit a distinct “bulbous” morphology, and excessive SQSTM1 accumulation could point to impaired autophagy, leading to toxic waste buildup, cellular stress, and synaptic dysfunction. These astrocytes have also been found to secrete MFGE8, with an emerging role as a linker between synapses and engulfing microglia^{44,45}. In our data, MFGE8 was up regulated but only in the PFC where there are the early signs of synapse loss.

This emphasises the tripartite synapse as a distinct cellular compartment differentially impacted by disease relative to the surrounding cellular environment. Notably, astrocytes locally translate transcripts within PAPs, a mechanism that may potentially become dysregulated in AD²⁶⁸. Finally, the astrocytic module featured DKK3, a Wnt antagonist elevated in AD that contributes to the loss of excitatory synapses²⁵⁶, likely through interactions with KREMEN1, a pathway known to exacerbate synapse loss when dysregulated²⁵⁵. These findings provide evidence for a synaptotoxic astrocytic signature associated with PAPs in AD. However, some proteins, such as Clusterin¹⁹⁴ and GPNMB²⁶⁹, have been associated with a protective response to synaptic damage, suggesting a partially protective response initiated by astrocytes in the peri-synaptic milieu. This is consistent with the recent observation that reactive astrocytes acquire both neuroprotective and deleterious signatures in response to Tau and A β pathology³⁹.

4.4.3 Synapse loss, synapse resilience and the role of PAPs

While direct measurement of synaptic maturity in *post mortem* tissue is not possible, synaptic maturity can be inferred from synaptic NMDA receptor subunit protein composition. Through synaptome mapping, Jessica Griffiths identified distinct populations of excitatory GluN1⁺ subtypes of synaptic puncta. The analysis revealed a loss of synapses and striking differential subtype vulnerability. Specific subtypes of excitatory synapses were maintained while others,

S2 and S4, were lost in the MTG of late-stage AD cases. Specifically, S2 and S4 subtypes with the highest relative GluN1⁺ protein concentration (i.e. fluorescent intensity), indicative of more mature synaptic puncta, and are more likely to be lost in AD. This loss is consistent with preclinical findings that report a reduced proportion of persistent dendritic spines²⁰⁷.

I identified a peri-synaptic astrocyte module associated with the loss of the same GluN1⁺ synapses subtypes, S2 and S4, particularly in the MTG region. Modelling S2 and S4 synaptic resilience showed that the greater astrocyte reactivity was associated with worse-than-expected synaptic resilience or greater synapse loss possibly by a greater synaptic engulfment or reduced stability of existing synapses. This could underlie the maladaptive astrocyte response where excessive astrocytic reactivity disrupts functions vital for synaptic maintenance. Supporting studies, such as a seminal study from Liddelow et al., revealed that neurotoxic A1 astrocytes lose synaptic-supportive functions, including those that promote neuronal survival, synaptogenesis, and phagocytosis of synaptic debris³². They showed that cultures of neurons exposed to A1 astrocytes displayed a 50% reduction in synapse density. Moreover, astrocytic dysfunction at the tripartite synapse may hinder synaptic plasticity and exacerbate loss, accelerating cognitive decline²⁷⁰ by disrupting potential compensatory mechanisms, such as synapse enlargement of surviving neighbouring synapses or regeneration²⁷¹. Interestingly, we also observed an increase in synapses with low GluN1 intensity, indicative of smaller synapses²⁷² and possible greater coverage by PAPs, which could lead to increased astrocytic uptake of glutamate. While this might help mitigate excitotoxicity, it could also impair efficient neurotransmission by reducing the available glutamate to spillover to neighbouring synapses^{273,274} or inhibit GABA release²⁷⁵.

4.4.4 Study limitations

A key strength of this study is the direct comparison of multiple brain regions, which allowed us to derive a spatiotemporal profile of synaptic changes over the disease progression timeframe. Furthermore, Jessica Griffiths employed synaptome mapping to identify distinct populations of excitatory synaptic puncta, enabling to uncover differential subtype vulnerabilities that might otherwise remain obscured in bulk analyses. However, the use of *post mortem* tissue and the cross-sectional design of this study limits its ability to fully capture the

true longitudinal course of disease progression. To mitigate this limitation, within-region case vs. control analyses were employed to account for regional variation and derived a pseudotime measure based on disease progression and Braak staging, enabling a proxy analysis of synaptic changes during disease progression. Furthermore, human tissue samples were obtained from later disease stages (Braak V-VI), leaving open the question of whether the remaining synapses, particularly in the MTG, where significant synapse loss was observed, are resilient or would eventually succumb to the same fate if disease progression had continued long enough. Based on the synaptic and accompanying astrocytic signatures, we propose that the remaining synapses are in the process of being lost (i.e. senescent) rather than exhibiting resilience. However, empirical testing is required to confirm this hypothesis.

The main technical limitation of this study is that SynPER synaptoneurosomes preparation method enriches for, rather than purifies, synapses. This approach allows for assessing the tripartite synapse, including astrocyte contributions, which is impossible with traditional synaptosome purifications. The synaptic fraction is, therefore, primarily of neuronal origin but may also contain peri-synaptic astrocyte processes (PAPs) and potentially other glial components present in close proximity to synapses. This presents a cell type of origin ambiguity: while proteins detected in this fraction are enriched for synaptic localisation, it is difficult to definitively determine whether a given protein originates from a neuronal synaptic compartment or from adjacent astrocytic or glial structures. This ambiguity complicates the interpretation of differential abundance, especially when proteins known to be expressed in multiple cell types (i.e. Clusterin) are detected. The possibility of contamination from non-synaptic components has also been highlighted by Hobson and Herzog²⁷⁶ following the Niu et al. paper²⁴⁸. To address this, we firstly characterised our synaptic proteome using a published synaptic protein exclusion list²⁴⁵ and secondly showed enrichment for PAP/astrocytic proteins in our astrocytic module within the synaptic fraction when considering all proteins. While this does not completely rule out astrocytic contamination, it provides evidence that astrocytic proteins located in the PAPs are altered in AD and are associated with synaptic pathology. Pilot validation staining highlighting the spatial proximity of GluN1 and Clusterin found in PAP further strengthened the evidence that astrocytic components closely interact with synaptic structures in AD (*Supplementary figure 4-9*). Further high-resolution imaging would be required to definitively resolve the precise localisation of Clusterin at the synapse and confirm

it originates from PAPs as current research presents conflicting views on its primary localisation, and protective or detrimental role^{194,277}.

In addition, sample size is a limitation of this analysis. The study includes synaptic and cytosolic proteomic profiling from 13 AD and 13 control donors across three brain regions (MTG, PFC, VC), limiting statistical power for detecting subtle or cell subtype-specific effects, particularly when stratifying across brain regions or correlating with clinical traits. Future studies with larger sample sizes and broader disease-stage coverage will be required to validate and extend these findings.

While synapse loss has been associated with cognitive decline in healthy aging, the relationship is often subtle and influenced by multiple confounding factors, including cognitive reserve and regional vulnerability²⁷⁸. In this study, age was controlled for to isolate AD-related effects; however, age-related synaptic changes could still contribute to variability in synapse density, especially in regions like the VC where AD pathology is minimal. Given the limited sample size, the analysis was underpowered to disentangle age-related from early disease-related synaptic alterations or detect more nuanced associations with cognitive trajectories in non-demented individuals. Larger, age-stratified cohorts will be essential to better understand how synapse loss contributes to cognitive decline during healthy aging versus neurodegeneration.

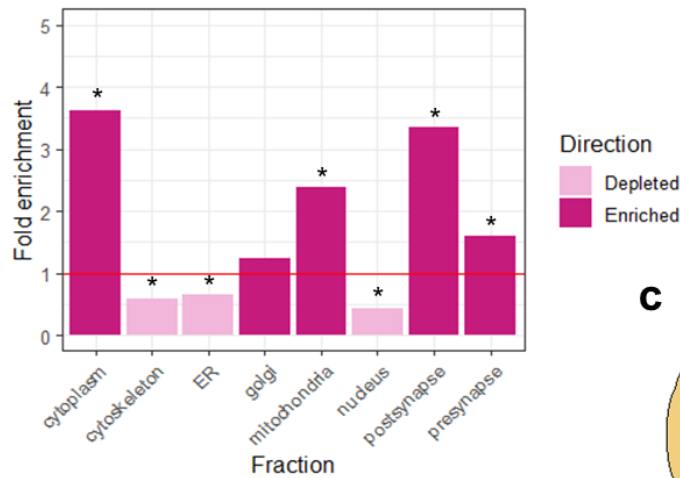
4.5 Conclusions and future work

In conclusion, this study provides substantial evidence for impaired synaptic health in the remaining synapses and highlights a neurotoxic peri-synaptic astrocyte (PAP) phenotype in AD. By building on previous research that characterised the synaptic proteome in AD, these findings provide novel insights into the time course and specificity of synaptic changes, and the role of astrocytes in synaptic pathology. This study opens new avenues for therapeutic interventions targeting astrocyte reactivity, potentially preserving synaptic resilience and alleviating cognitive impairment in AD. Future work should prioritise the validation of this spatiotemporal model of synaptic changes against Braak staging, by including samples from all the Braak spectrum, specifically mid-stages III-IV, which were missing from this study. Secondly, *in vitro* studies should investigate the detrimental role of astrocytes at vulnerable

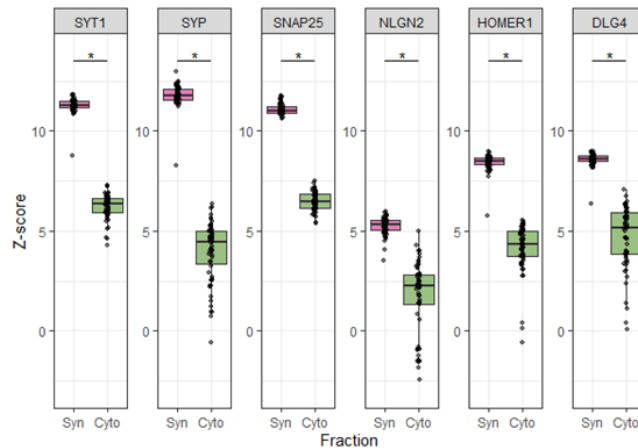
synapses and how these may interfere with synaptic plasticity and exacerbate the loss of certain subtypes. For instance, co-culture systems involving neurons and astrocytes could be developed to model tripartite synaptic interactions in controlled settings. These systems could then be perturbed, by pharmacologically targeting PAPs, interfering with ligand-receptor interactions, or genetically altering astrocytic reactivity, to evaluate the effects on synaptic stability, plasticity, and subtype-specific resilience. The ligand and receptor candidates highlighted in the NicheNet analysis could be a starting point for this. Finally, future studies should also explore other synapse types, such as inhibitory synapses, to determine whether similar proteomic and spatial distribution patterns occur across different synaptic subtypes.

4.6 Supplementary figures

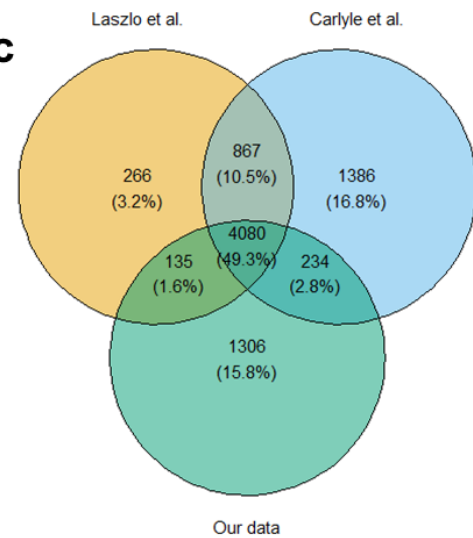
a



b



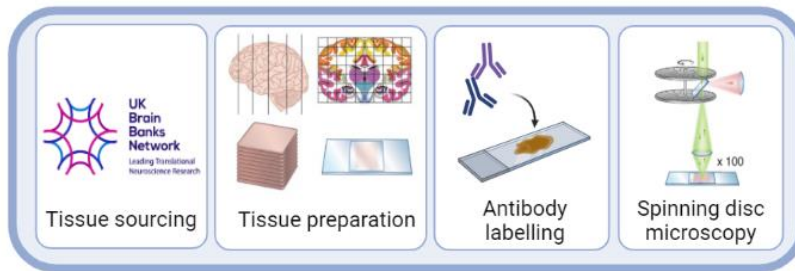
c



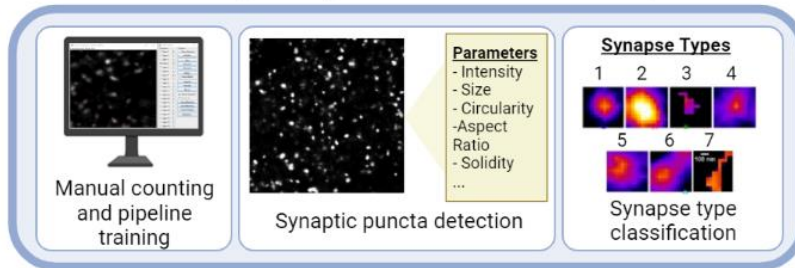
Supplementary figure 4-1 Synapse enrichment in human post-mortem cortical tissue

a) Comparison of detected proteins in the synaptic fraction against a consensus list of unique cellular fraction-associated protein IDs (*Fisher's test, Bonferroni-adjusted p -value < 0.001). **b)** Boxplots of the normalised abundance of selected pre- (SYT1, SYP, SNAP25) and post-synaptic (NLGN2, HOMER1, DLG4) proteins in synaptic fraction (Syn) compared to cytosolic (Cyto) (* $p \leq 0.05$, Wilcoxon rank sum/Mann-Whitney U test with continuity correction). **c)** Venn diagram illustrating the intersection of 5755 uniquely identified proteins with previous human synapse proteomic datasets.

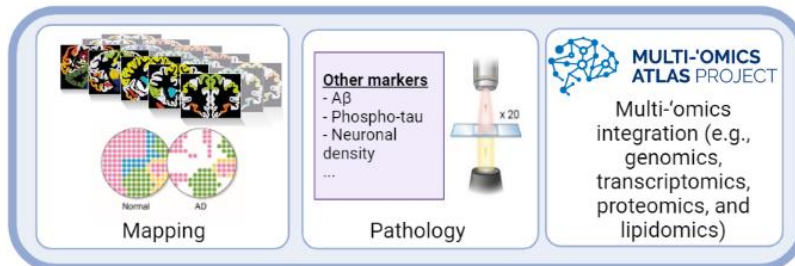
1. Synapse labelling and imaging



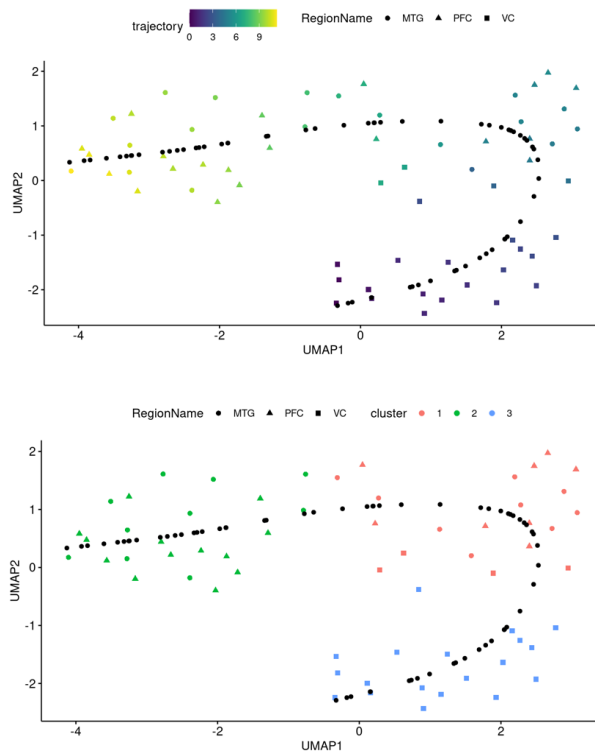
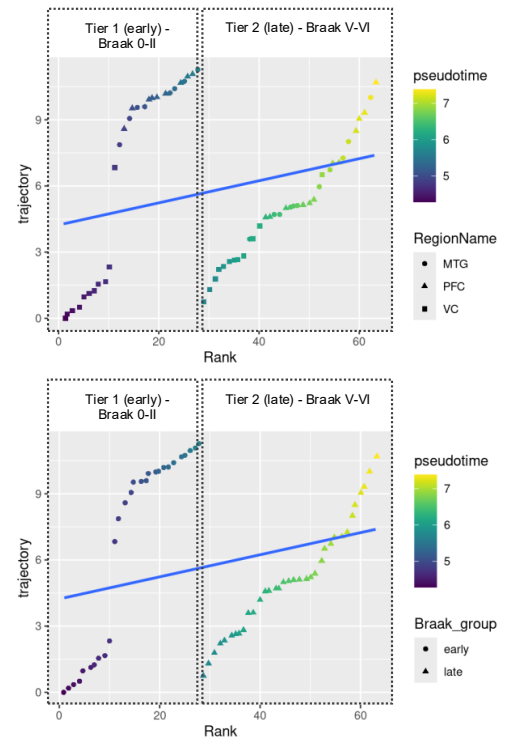
2. Image analysis and machine learning



3. AD synaptome mapping and data integration



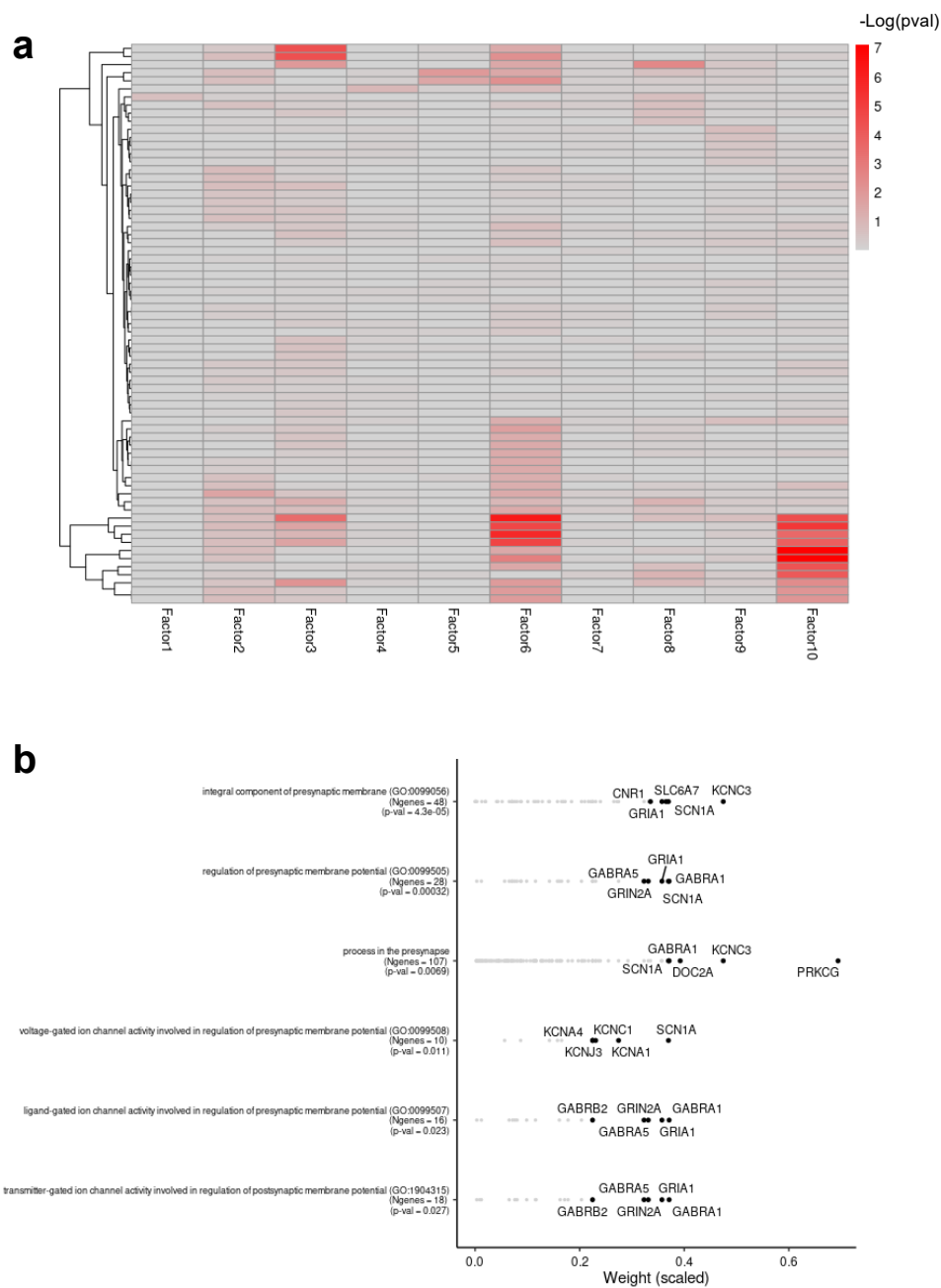
Supplementary figure 4-2 Synaptome mapping graphical abstract

a**b**

Supplementary figure 4-3 Semi-supervised approach for pseudotime derivation

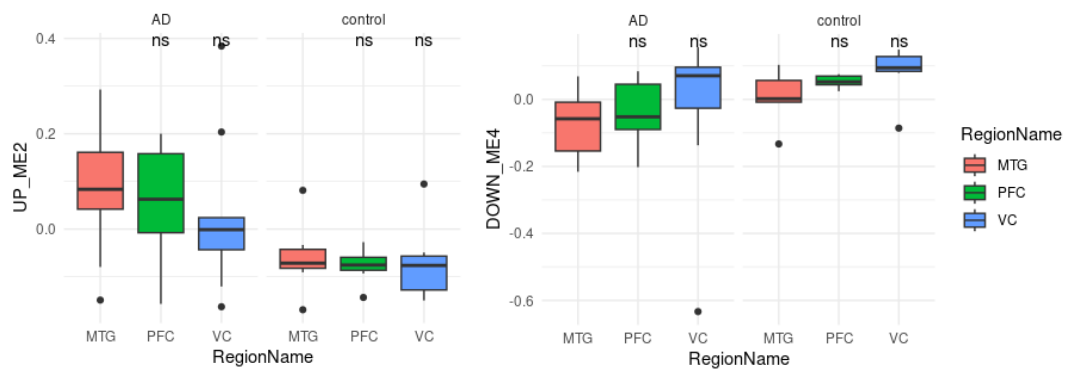
a) AD-relevant factors 1,2, and 3 were used in a semi-supervised approach to derive a pseudotime measurement. This figure presents a Uniform Manifold Approximation and Projection (UMAP) plot illustrating the distribution of brain tissue samples based on inferred pseudotime (i.e. trajectory) derived via Slingshot. Each point represents a tissue sample coloured by inferred pseudotime, ranging from early (blue) to late (yellow) stages of the disease. The VC dense region (cluster 3) was chosen as the root of the slingshot trajectory.

b) This plot illustrates the inferred trajectory of AD progression plotted against the ranked positioning of brain tissue samples according to Braak staging criteria into tiers. The x-axis represents the overall rank of samples from early to late Braak stages, while the y-axis denotes the trajectory inferred through pseudo time analysis. The linear model depicts the relationship between sample ranking within the tiers and disease trajectory. Fitted values correspond to the final pseudotime value.



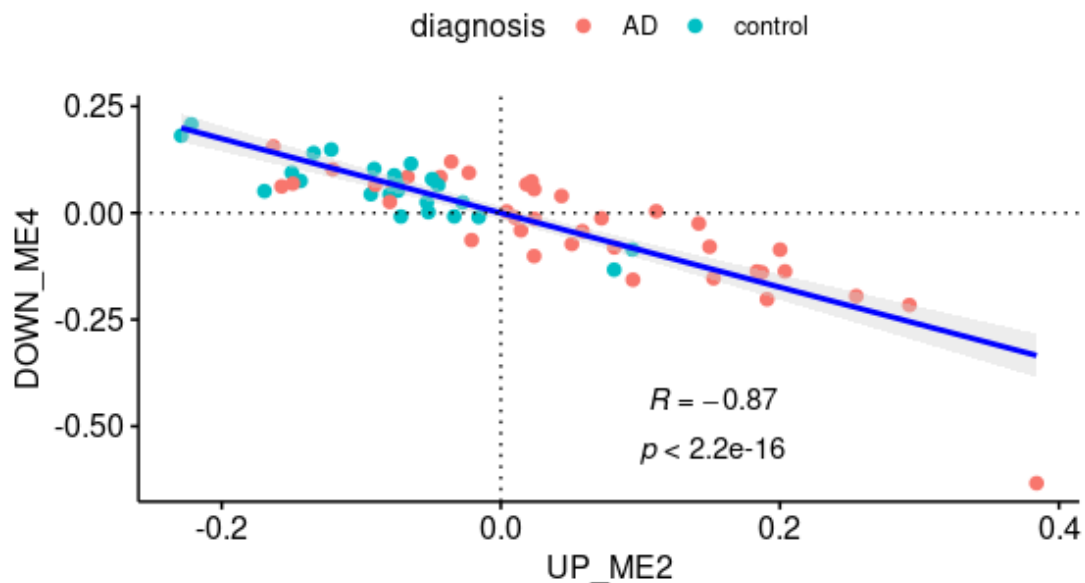
Supplementary figure 4-4 MOFA factor SynGO enrichments

a) Heatmap with the adjusted $-\log(p\text{-values})$ that result from the feature set enrichment analysis. Rows are SynGO feature sets, and columns are factors. **b)** Each row corresponds to a significant pathway from factor 3, sorted by statistical significance, and each dot corresponds to a protein. The top 5 proteins sorted by the weight in factor 3 are displayed for each pathway. The remaining proteins are coloured in grey.



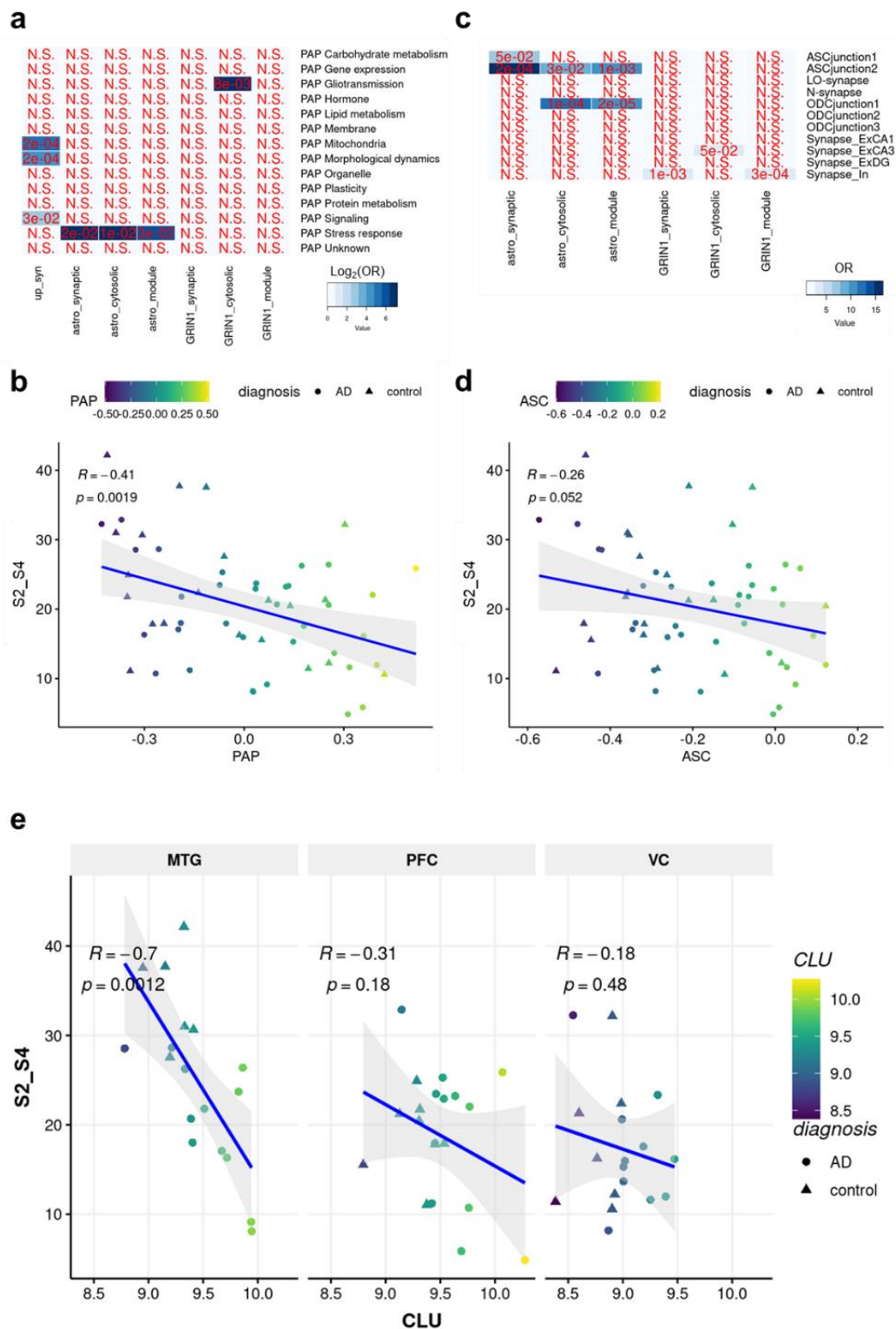
Supplementary figure 4-5 Prioritised module eigenvalues regional differences stratified by case-control status

Boxplot of module eigenvalue across brain regions, with statistical comparisons performed using a paired t-test. The reference group is MTG, and p-values indicate significance levels. Paired samples were matched by case ID across all three regions.



Supplementary figure 4-6 Inverse correlation between synaptic and astrocyte modules

Linear regression of F3_DOWN_ME4 eigenvalues (Synaptic module) against F3_UP_ME2 eigenvalues (Astrocyte module).

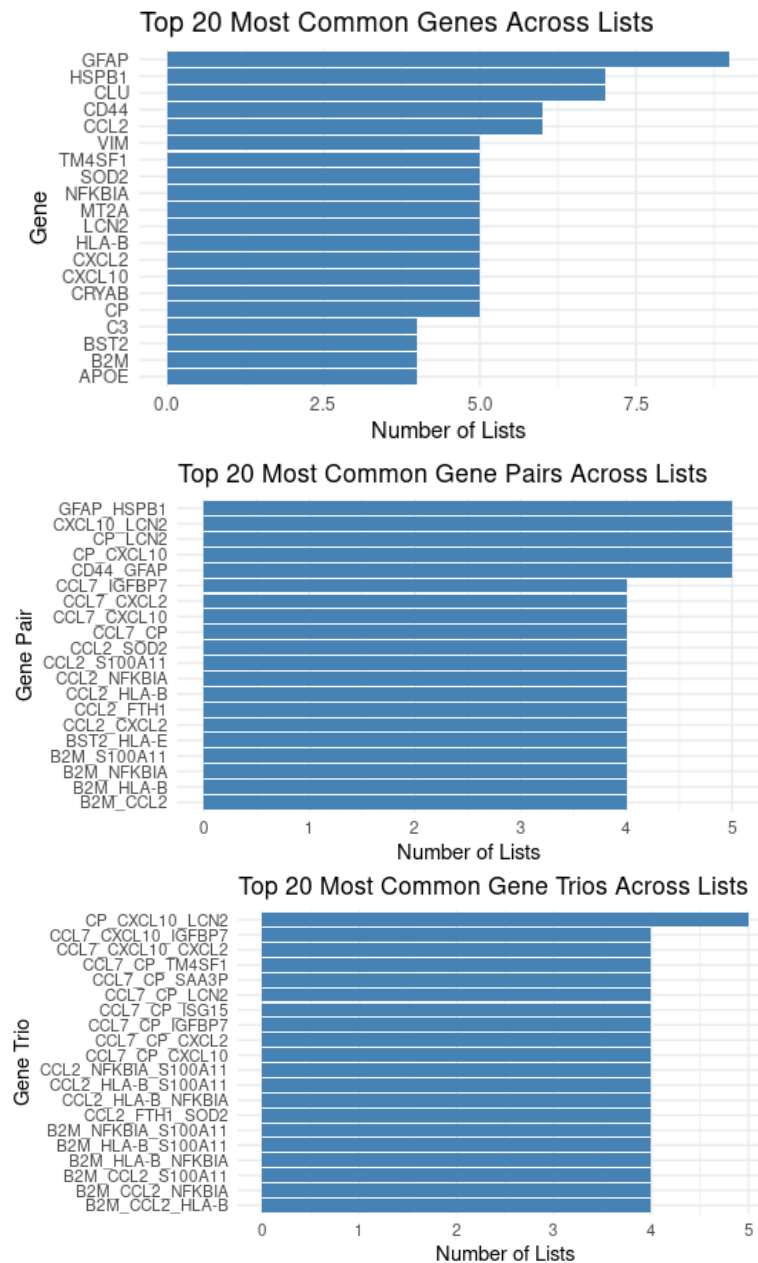


Supplementary figure 4-7 PAP & ASC gene set activity

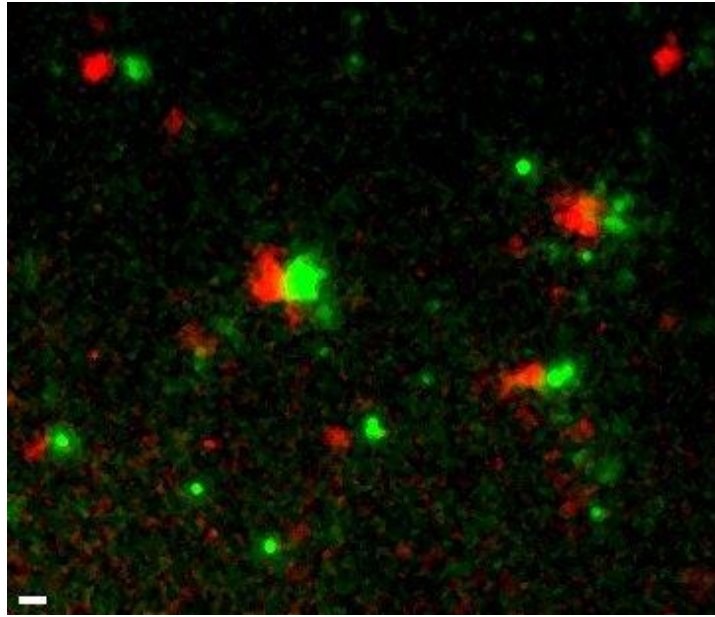
a) Heatmap of pairwise enrichment analysis of (upregulated proteins in synaptic fraction, ME2 astrocyte module, ME4 GluN1⁺ module) against PAP gene set. Fisher's exact test determines the *p*-value and odds ratio compared to a genomic background (Colour = log₂(OR), *p*-value).

b) Linear regression of the summed density of S2 and S4 synapse subtypes (S2_S4) against PAP activity (via GSVA) in the synaptic proteomics fraction. **c)** Heatmap of pairwise

enrichment analysis of (upregulated proteins in synaptic fraction, ME2 astrocyte module, ME4 GluN1⁺ module) against ASC gene set. Fisher's exact test determines the p-value and odds ratio compared to a genomic background (Colour= OR, p-value). **d)** Linear regression of the summed density of S2 and S4 synapse subtypes (S2_S4) against ASC activity (via GSVA) in the synaptic proteomics fraction. **(e)** Linear regression of the summed density of S2 and S4 synapse subtypes (S2_S4) against CLU abundance in the synaptic proteomics fraction.



Supplementary figure 4-8 Literature reactive astrocytes signatures: 20 most common single, pair and gene trio
Lists included are available in Supplementary Table 10.



Supplementary figure 4-9 *In vitro validation of Clusterin colocalisation with synaptoneurosomes*

Red: GluN1⁺ puncta, Green: Clusterin. Scale bar= 1μm

4.7 Supplementary tables

https://eleonoreschneegans.shinyapps.io/supplementary_4/

Supplementary Table 4-1	PAP/ASC list
Supplementary Table 4-2	Carlyle synaptic inclusion list
Supplementary Table 4-3	Differential expression AD vs. Control
Supplementary Table 4-4	PHF1/amyloid regression
Supplementary Table 4-5	Subject Demographics
Supplementary Table 4-6	SynGo Pathways
Supplementary Table 4-7	GO biological pathways
Supplementary Table 4-8	MOFA variance
Supplementary Table 4-9	Module protein memberships
Supplementary Table 4-10	Literature reactive astrocyte
Supplementary Table 4-11	Summary of thesis reactive astrocyte signatures
Supplementary Table 4-12	Astrocyte pathways contributing to synaptic dysfunction

Chapter 5 Spatial omic characterisation of glial activation in Alzheimer's disease using imaging mass cytometry

Chapter declarations:

1. Some of the work in this chapter has been previously made available as a preprint:

Fancy, Nurun, Nanet Willumsen, Stergios Tsartsalis, Combiz Khozoie, Aisling McGarry, Robert Muirhead, **El onore Schneegans**, Karen Davey, Vicky Mn Chau, Amy M. Smith, William Scotton, John Hardy, Dann Huh, Paul M. Matthews and Johanna S. Jackson (2023). "Mechanisms contributing to differential genetic risks for TREM2 R47H and R62H variants in Alzheimer's Disease". In: *medRxiv*. doi: [10.1101/2022.07.12.22277509](https://doi.org/10.1101/2022.07.12.22277509).

[CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/)

2. I thank the following contributors:

- Wenqi Cho: Tested and optimised the Steinbock pipeline on an in-house test IMC dataset (MSc Summer student supervised by Johanna Jackson and me. Thesis: single-cell spatial omics analysis of AD using IMC, 2023).

All the internal data used for analysis was generated entirely by the wet lab team, specifically:

- Vicky Chau (PhD student in the Matthews' group): IMC data generation, plaque segmentation, spillover correction.

Abstract

Background: AD is marked by the accumulation of amyloid-beta plaques and neurofibrillary tangles, leading to neurodegeneration. Understanding the spatial heterogeneity of glial cell activation and their interactions within the plaque microenvironment is crucial for characterising local cellular response to AD pathology. Imaging Mass Cytometry (IMC) enables high-dimensional, spatially resolved protein analysis at the single-cell level, offering detailed mapping of cellular phenotypes and interactions.

Objectives: This study aimed to validate the glial subpopulation identified in *Chapter 3* at the spatial level and to characterise the spatial distribution of activation states of microglia and astrocytes in relation to plaque niches in *post mortem* AD brains.

Methods: Brain tissue samples from the middle temporal gyrus and somatosensory cortex of 43 AD donors were analysed using IMC with a panel of 40+ metal-conjugated antibodies targeting cell-specific and functional markers. Downstream analyses identified glial cell subpopulations and their spatial distribution relative to amyloid plaques, classified as non-neuritic (4G8⁺/AT8⁻) or neuritic (4G8⁺/AT8⁺). Cells were assigned to their respective plaque niche defined as those located on the plaque or within 10 µm of its edge. Pseudotemporal analysis and cell-cell interaction analyses were employed to assess changes in cellular dynamics within plaque niches.

Results: I identified two subtypes of activated microglia: “C1q⁺” and “HLA-DR⁺” microglia. Additionally, I identified two astrocyte subpopulations: (1) “GFAP/S100β⁺” classical reactive astrocytes, associated with an overall net neuroprotective effect, and (2) “Neuro/Astro” astrocytes localised near neurons, characterised by mixed intermediate reactive and exhausted features, contributing to a net neurotoxic effect. Plaque niches were enriched with activated microglia and neurotoxic mixed reactive/exhausted astrocytes, with protein expression within these cells significantly influenced by proximity to plaques. Reduced microglial activation and a decreased proportion of GFAP/S100β⁺ reactive astrocytes were observed in *TREM2*^{R62H/R47H} carriers, along with increased neurotoxic mixed reactive/exhausted astrocytes (“Neuro/Astro”). Pseudotemporal analysis revealed an imbalance toward neurotoxic astrocyte exhausted states and altered cell-cell interactions, highlighting a shift towards a harmful maladaptive glial response as plaques evolved.

Conclusions: The study demonstrates the utility of IMC in spatially resolving glial cell subpopulations and how imbalances in the local plaque microenvironment may contribute to disease progression.

Graphical abstract

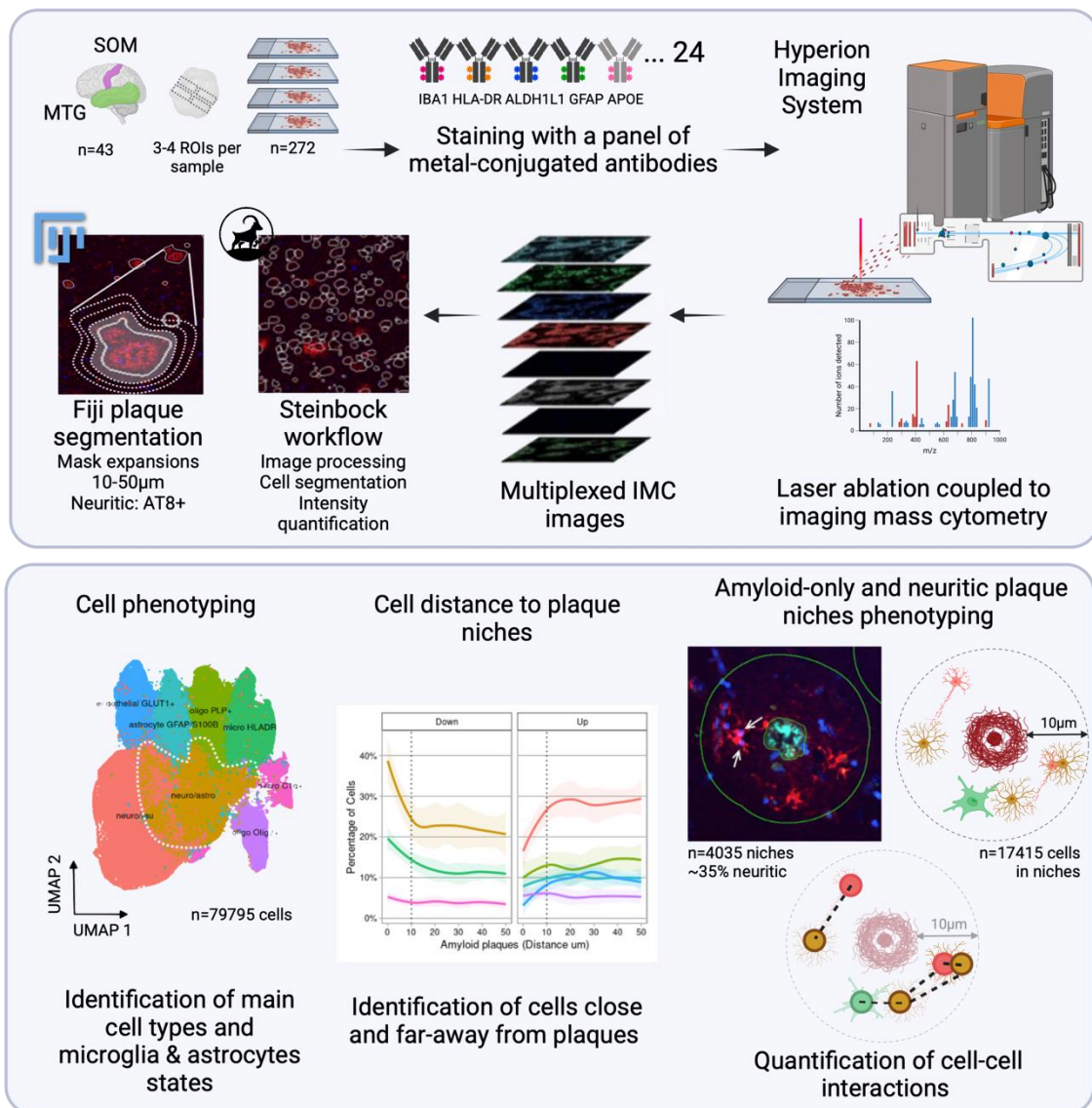


Figure 5-1 Chapter 5 Graphical abstract
Figure generated via [Biorender](#).

5.1 Introduction

5.1.1 Spatial omics: integrating molecular profiling into a spatial context

The emergence of spatial omics has transformed the study of disease biology by improving our understanding of tissue microenvironments, the spatial variability of gene expression and protein profiles, and revealing changes in cell-cell interactions. This has proven particularly useful in the oncology field, where spatial omics have dramatically improved our ability to analyse tumour microenvironments and patients' responses to treatment based on spatial analyses of tissue biopsies^{279,280}. However, spatial technologies are not limited to cancer — they are becoming increasingly important in studying neurodegenerative diseases like AD. Although the use of spatial omics for brain research is not destined to mirror the tissue biopsy methods used in cancer, it nevertheless holds promise for characterising localised molecular disease mechanisms on *post mortem* tissue samples and will ultimately inform on new treatment strategies²⁸¹.

While bulk omics approaches have led to substantial improvements in our understanding of AD, they are limited by their inability to capture spatial heterogeneity and cell-cell interactions — key to unlocking the brain's pathological microenvironment. Although traditional immunohistochemistry (IHC) and immunofluorescence (IF) have been used to map proteins within tissues, more recent spatial omic techniques now allow higher-throughput profiling of transcripts and proteins with spatial resolution. Spatial molecular profiling methods, such as GeoMx and Visium, combine sequencing or mass spectrometry with spatial information from tissues (via spatial barcoding or microdissection-based methods). The other prominent family of spatial omic methods is molecular imaging, encompassing Imaging Mass Cytometry (IMC), CODEX, CyCIF, MERFISH, and more. These directly visualise molecules labelled by antibodies or probes via imaging modalities. Regardless of the method, all spatial omics technologies fall into these four steps: “**Detect** [Antibodies, DNA probes, barcoded probes], **Identify** [Next Generation Sequencing, Mass Spectrometry, hybridisation signals of labelled probes], **Measure** [Microscopy, Sequencing] and **Locate** [Coordinates, barcoding]”²⁸².

This study employs IMC using the Hyperion imaging system, which integrates TOF (Time of Flight) cytometry with immunohistochemical staining and imaging to analyse up to 40 protein markers in a single tissue section. The presence and location of target protein markers are indicated by detecting and measuring signals from antibodies labelled with metal isotopes. This allows for tissue characterisation at the single-cell level. I use data from IMC to examine variations in protein expression and localisation relative to plaque niches, as well as cell-cell interaction changes between relevant cell types/cell subpopulations of interest. IMC has previously been used for analysing single-cell phenotypes, their functional states and cell-cell interactions in multiple sclerosis post-mortem human tissue²⁸³, though has not been used to study AD other than by our group (Matthew's group).

5.1.2 Studying AD from a spatial angle

Several recent studies have leveraged spatial omics technologies to investigate AD. To begin with, regarding the AD animal model category, Chen et al.²⁸⁴ used spatial transcriptomics to identify a plaque-induced gene (PIG) network associated with microglial and astrocytic responses to amyloid plaques, uncovering spatial variability in glial activation states. Similarly, Mallach and colleagues evidenced that amyloid plaque niches are microglia-dense and induce a neurotoxic astrocyte phenotype responsible for disrupting glutamatergic synaptic signalling in hippocampal neurons⁴³. In addition, Choi et al. used spatial transcriptomics to track glial activation across disease stages in a 5xFAD AD mouse model, identifying spatial associations between genes such as *GFAP* and *TREM2* with amyloid plaques²⁸⁵. More recent work by Johnston et al. highlights how spatial transcriptomics can reveal distinct microglial and astrocytic activation patterns with plaque proximity in the context of the *TREM2* mutation²⁸⁶. Such spatial analyses underscore the importance of studying glial-astrocytic interactions in situ to unravel how these cells contribute to neurodegeneration in AD.

Fewer studies have employed human post-mortem brain samples to investigate spatial dynamics in AD. A seminal paper from Chen et al. performed spatial transcriptomics using Visium on *post mortem* AD and control samples and revealed genes associated with the vulnerability of the MTG²⁸⁷. A recent preprint performed spatial transcriptomics on

postmortem tissue from 21 ROSMAP donors and found multiple glial states around amyloid plaques, with aberrant synaptic and neuronal loss in these “high-glia” areas²⁸⁸. Caramello et al. used IMC to identify neuronal subtypes vulnerable (RORB⁺) to early degeneration at the protein level, and accumulating intracellular A β ¹⁸⁵. Their findings suggest that reactive microglia are spatially associated with these vulnerable neurons. Similarly, Jury-Garfe et al. used spatial transcriptomics study and found that in the brains of asymptomatic AD donors, the amyloid plaque microenvironment includes microglia with enhanced motility and a reduced capacity for tau seeding compared to those from symptomatic AD donors, pointing to potential mechanisms of cognitive resilience²⁸⁹. Finally, Muñoz Castro et al. utilised a cyclic IF workflow and identified three distinct states of microglia and astrocytes in AD brains (homeostatic, intermediate and reactive). This highlights the potential for studying glial states with few markers, provided that the spatial organisation is taken into account²⁹⁰.

5.1.3 The plaque niche: AD’s neuropathological microenvironment

AD is characterised by the accumulation of aggregated A β peptides in senile plaques. A β peptides originate from amyloid precursor proteins (APP), produced via the sequential proteolytic cleavage of APP by β -secretases and γ -secretases, leading to various A β species. A β can aggregate into various assemblies, ranging from small soluble oligomers to larger insoluble fibrils, ultimately forming amyloid plaques in the brain^{24,291}. Some forms of the aggregated peptides, such as A β 42, are toxic to neurons. Amyloid plaques are classified into diffuse and dense-cored, where a subset of dense-core plaques are neuritic plaques (NP)²⁹². The 4G8 marker, which detects amino acids 17-24 of the amyloid- β peptide, is widely used to identify amyloid plaques²⁹³. In contrast, neuritic plaques, which consist of amyloid cores surrounded by dystrophic neurites, are tau⁺ and associated with more advanced neuronal damage and cognitive impairment²⁹⁴. This study uses the AT8 marker to detect hyperphosphorylated tau — a hallmark of neurofibrillary tangles and neuritic plaques — and an ideal marker for detecting early tau pathology (i.e. as opposed to PHF1 that is more specific to mature tau aggregates)²⁹⁵. Although the precise trigger for the change from diffuse to neuritic remains unclear, it is accepted that NP are developing from diffuse plaques, reflecting an evolution in pathology severity²⁹⁶.

The concept of plaque niche is constantly evolving but had been observed initially: “The presence of reactive microglial cells and astrocytes in β -amyloid plaques of Alzheimer's disease has already been observed by Alzheimer himself”²⁹⁷. Pivotal observations by McGeer et al. further established the close association of microglia with senile plaques²⁹⁸. Heneka et al., in the early 2000s, explored the role of glial activation near amyloid plaques, leading to the recognition of the spatially restricted influence that plaques exert on surrounding cells²⁹⁹. The concept was refined in recent years as spatial omics techniques developed. The term “plaque niche” appeared in a landmark 2020 paper²⁸⁴, which further characterised the cellular phase of AD¹⁸, now at a spatial level. Ever since, there has been a consensus that a plaque niche refers to a microenvironment surrounding an amyloid plaque in brain tissue, typically within a defined spatial radius (often 10-50 μm ^{43,284}), where cells such as astrocytes, microglia, neurons, and other glial cells are influenced by the presence of the plaque and may exhibit dysfunctional behaviours ranging from detrimental inflammatory response, defective clearance, lipid metabolism leading to altered synaptic signalling and neuronal death¹⁸.

5.1.4 Aim and objectives

The overarching aim of this chapter is to characterise the activation states of microglia and astrocytes within the neuropathological microenvironment of AD at a spatial proteomic level, using IMC. The hypothesis is that microglial and astrocyte activation states vary spatially across distinct pathological niches, such as non-neuritic and neuritic plaques, and that these activation patterns change with disease progression.

To address this hypothesis, the following objectives have been set:

1. To validate microglia and astrocyte subtypes identified through snRNA-seq at the spatial level using IMC.
2. To perform molecular and cellular footprinting of non-neuritic and neuritic plaque niches in AD and characterise the spatial variability of glial activation states in relation to plaques.
3. To explore the evolution of cell-cell interactions within the plaque microenvironment.

5.2 Methods

5.2.1 Imaging Mass Cytometry

This sub-section of the methods was performed and written by the wet lab team but is included because it is essential for understanding the data generated and used in this chapter. These methods are also found in publications from the group^{31,165}.

FFPE 5–10 μm sections from 43 AD cases, including tissue from both the MTG and SOM were used. Before ablation, sections were immunostained using lanthanide-tagged antibodies (**Table 5-1**). The antibody panel was designed to comprehensively profile glial–neuronal interactions and cellular states relevant to AD pathology across brain regions. Markers were selected to capture (1) major brain cell types, including neurons (NeuN, MAP2), astrocytes (GFAP, S100 β , ALDH1L1), oligodendrocytes (PLP, OLIG2), and microglia/macrophages (Iba1, CD68, MS4A4A, P2Y12, HLA-DR, CD163, CSF1R, CD206, P2Y12); (2) reactive or disease-associated states linked to lysosomal activity (LAMP1), lipid metabolism (ApoE), inflammation (IP-10), complement (C1q); and (3) AD-relevant neuropathology, including amyloid plaques (4G8, APP) and tau (PHF1, AT8).

The slides underwent routine dewaxing and rehydration before undergoing antigen retrieval in a pH8 Ethylenediaminetetraacetic acid (EDTA) buffer. The slides were then treated with a 10% normal horse serum (Vector Laboratories) blocking solution before incubation with an antibody cocktail at 4°C overnight. The slides were washed in 0.02% Triton X-100 (Sigma-Aldrich) before incubation with the Iridium-intercalator (Fluidigm) and then washed and air-dried. All antibody conjugation was performed using the Maxpar X8 protocol (Fluidigm).

IMC was performed using a Hyperion Tissue Imager (Standard BioTools, San Francisco, USA) coupled to a Helios mass cytometer. The instrument was first tuned utilising the manufacturer's 3-Element Full Coverage Tuning Slide before loading the slides into the device. Four 500x500 μm^2 regions of interest (ROI) within the grey matter were selected and ablated using a laser operating at a frequency of 200Hz with 1 μm resolution. The data was stored as .mcd files compatible with MCD Viewer software (Fluidigm).

<i>Antibody</i>	<i>Gene name</i>	<i>Supplier</i>	<i>Metal isotopes</i>	<i>dilution (1:x)</i>
<i>PLP</i>	PLP1	Invitrogen	141Pr	1000
<i>GFAP</i>	GFAP	Fluidigm	143Nd	500
<i>4G8</i>		Biologend	144Nd	200
<i>NeuN</i>	RBFOX3	Merck	145Nd	30
<i>S100β</i>	S100β	Fluidigm	146Nd	2000
<i>CD163</i>	CD163	Fluidigm	147Sm	200
<i>MAP2</i>	MAP2	Fluidigm	148Nd	400
<i>CSF1R</i>	CSF1R	Abcam	150Nd	400
<i>LAMP1</i>	LAMP1	Fluidigm	151Eu	100
<i>CD206</i>	CD206	Fluidigm	153Eu	200
<i>APP</i>	APP	Thermofisher	154Sm	50
<i>MS4A4A</i>	MS4A4A	Abcam	155Gd	100
<i>PHF1</i>		Peter Davies	156Gd	500
<i>C1q</i>	C1QA	Abcam	158Gd	350
<i>CD68</i>	CD68	Fluidigm	159Tb	200
<i>ApoE</i>	APOE	Novus	160Gd	100
<i>ALDH1L1</i>	ALDH1L1	Thermofisher	162Dy	50
<i>AT8</i>	MAPT	Thermofisher	165Ho	400
<i>IP-10</i>	CXCL10	Bio-Techne	168Er	50
<i>Iba1</i>	AIF1	WAKO	169Tm	500
<i>OLIG2</i>	OLIG2	Abcam	171Yb	50
<i>P2Y12</i>	P2RY12	Abcam	172Yb	50
<i>HLA-DR</i>	HLA-DR	Fluidigm	174Yb	200
<i>GLUT1</i>	SLC2A1	Abcam	176Yb	16000

Table 5-1 IMC antibody panel

Methods in sections 5.2.1 to 5.2.5 were first tested and optimised by Wenqi Cho (MSc Summer student) and me on another in-house IMC dataset and then formally applied to this chapter's dataset by Vicky Chau (PhD student in the Matthews' group).

5.2.2 Steinbock workflow & cell segmentation

Image pre-processing steps from image conversion from .mcd to TIFF files, hot pixel filtering, segmentation and intensity measurements were performed using the Steinbock pipeline (v0.16.1)³⁰⁰. Single-cell segmentation was performed via the DeepCell algorithm integrated within Steinbock, a pre-trained deep-learning segmentation algorithm trained on TissueNet³⁰¹. Single cells were segmented based on the nuclear channel via the command *Steinbock segment*

deepcell --minmax --type nuclear. Workflow outputs, including images, single-cell channel intensities, and spatial object properties, can be loaded and analysed later in R.

5.2.3 Plaque segmentation

A custom macro was employed to process the IMC images on the Fiji software (v2.9.0)³⁰². The macro applied relevant image transformations, including contrast enhancement, despeckling, and Gaussian blurring, followed by Otsu thresholding to segment amyloid plaques (4G8) and tau pathology (AT8). Plaques were classified as either 4G8⁺ AT8⁺ neuritic or 4G8⁺ AT8⁻ amyloid-only (i.e. non-neuritic) plaques. Plaque masks were enlarged from the plaque edge at various pixel distances (10, 20, 30, 40, and 50 pixels, where 1 pixel = 1µm).

5.2.4 Cell distances

Cell masks were matched to the plaque masks using the Steinbock function *Steinbock utils match_cell_masks plaque_masks -o matched_objects*, allowing for the identification of overlapping cells and plaques at various distances. Each cell was assigned a proximity distance to the closest plaque (on plaque, 10, 20, 30, 40, 50), with the smallest distance from the nearest plaque mask expansion being used for cells overlapping multiple expanded plaque masks. For exact cell distances, I used the *minDistToCells()* from the *imcRtools* package (v1.10.0)³⁰³ to assign the distance to the closest cell on a plaque, specifying cells that are on plaque for the *x_cell* parameter, thereby allocating an exact distance for all the cells not found to be on plaques.

5.2.5 IMC data processing

The resulting data were then imported into R for further analysis using the *imcRtools* (v1.10.0)³⁰³ package and loaded onto a *SpatialExperimentObject*³⁰⁴ via the *read_steinbock()* function. Single-cell expression data were transformed using the hyperbolic arcsine function (*arcsinh*) with one cofactor, a best-practice transformation to stabilise variance and normalise data with a wide dynamic range, which is particularly useful for IMC data containing many zero values.

$$\text{asinh}(x) = \log(x + \sqrt{x^2 + 1})$$

Images and single-cell intensities were corrected for spillover using a spillover matrix estimated from single metal spot staining as suggested by Chevrier et al.³⁰⁵ by my colleague Vicky Chau. Each antibody conjugated to a metal isotope was dotted separately on an agarose gel and analysed across all isotopes with the Hyperion system. In R, the spillover matrix was then estimated using the *computeSpillmat()* from the Catalyst package, quantifying how much of the signal from a specific metal isotope contaminates the measurements in neighbouring channels due to isotopic mass overlap or detector sensitivities (**Supplementary figure 5-1**). Then, spillover compensation on the single-cell mean intensities per channel was applied using the *compCytotf()* function.

Cells were retained to exclude cells with atypically large or small sizes if their area fell within a defined interquartile range (IQR), specifically greater than $Q1 - 1.5 \times IQR$ and less than $Q3 + 1.5 \times IQR$. ROIs with cell counts in the top and bottom 1% of the distribution were excluded to mitigate the impact of outliers.

The remaining methods described in this chapter were designed and carried out by me.

Control images were filtered out to focus on AD-only images. I used the *Seurat* package (v4.3.0)³⁰⁶ to integrate data across MTG and SOM brain regions to correct batch effects and harmonise potential brain region-related differences. First, the data was split by brain region using *SplitObject()*. Each region was scaled with *ScaleData()* and subjected to PCA. Integration anchors were identified with *FindIntegrationAnchors()*, using the "rpca" method and the first 12 dimensions. Finally, I used *IntegrateData()* to merge the datasets, relying on the first five principal components for integration.

5.2.6 Cell phenotyping

After integrating the data across brain regions, I performed cell clustering using the *Seurat* package³⁰⁶. First, the integrated data was scaled using *ScaleData()*, followed by a PCA with 30 components [*RunPCA()*] to reduce dimensionality. Based on the elbow plot [*ElbowPlot()*], I selected the top 10 PCs for downstream analysis. I then constructed a shared

nearest neighbour (SNN) graph using *FindNeighbors()* and identified distinct cell clusters with *FindClusters()* at a resolution of 0.1. The resulting clusters were visualised on a UMAP and assigned to major cell types based on marker expression in individual clusters. Astrocytes were identified by high expression of ADH1L1, GFAP and S100 β . Markers, including C1q and HLA-DR, identified microglia. Oligodendrocytes were identified by the markers OLIG2 and PLP. The marker GLUT1 identified endothelial cells. Neurons were identified by the neuronal markers MAP2 and NeuN. Based on marker expression, astrocytes and microglia clusters were further assigned to subpopulations or glial states.

5.2.7 Correspondence IMC-snRNA

I implemented several steps to assess the correspondence between IMC and snRNA-seq data. First, I computed module scores for each cell using the Seurat *AddModuleScore()* function. The features chosen for these scores were specific to the studied glial subtype, as indicated in **Table 5-2**. Then, these scores were normalised to a 0-1 range using min-max scaling to ensure comparability across the two modalities. I then calculated the difference between the scaled snRNA and IMC scores for each sample as a measure of correspondence where a difference close to zero is expected for similar scores. To evaluate whether the score differences were significantly different from zero, a one-sample t-test was performed for each subcluster, with the null hypothesis that the mean difference was zero. I grouped the data by subclusters and tested the null hypothesis for each, using the results to interpret significance. If the p-value was below 0.05, it indicated poor correspondence between IMC and RNA scores for that subcluster (i.e. because of a significant difference). Still, if not significant, it suggested good correspondence between the two data types.

<i>IMC cluster</i>	<i>IMC marker</i>	<i>snRNA marker</i>
<i>Micro HLA-DR⁺</i>	IBA1; HLA-DR	AIF1; HLA-DR
<i>Micro C1q⁺</i>	C1Q	C1Q1A; CQ1B; CQ1C
<i>Neuro/Astro</i>	ALDH1L1; APOE; MAP2	ALDH1L1; APOE; MAP2

Table 5-2 Markers chosen to compute IMC-snRNA correspondence scores

5.2.8 Cell type proportion tests via permutations

Using the *permutation_test()* function from the *scProportionTest* package (v1.0.0)³⁰⁷, I assessed the relative proportions of cell types within plaque niches compared to non-niche cells at an FDR threshold of 0.05, where cells on plaques and within 10 μm are considered to be in the niche. This permutation test compares the observed proportions of each cell type within plaque niches to those in non-plaque-niches cells by generating a distribution of differences through 1000 random permutations of the sample labels. This analysis was stratified by plaque type (amyloid-only [non-neuritic], neuritic and all plaques), and the results were visualised using the *ggforestplot* (v0.1.0) package. The same analysis assessed differences in the relative proportion of niche-associated cell types for *TREM2* variants relative to common variant (CV). Ultimately, I repeated this approach to assess differences in cell types predominantly found outside of niches between *TREM2* variants and CV.

5.2.9 Differential protein expression analysis relative to distance from plaque distance

I used the *zlm()* function from *MAST* package (v1.20.0)¹⁷³ with the glmer method to fit a generalized linear mixed model to model marker expression as a function of cell distance to plaque while accounting for sample-specific random effects [*marker* ~ *distance* + (*I*|*sample_id*)]. All cells on plaques were set to a distance of 0. The *zlm()* function assesses log-fold changes for each marker and performs likelihood ratio tests (LRT) to evaluate the significance of the *distance* effect. The LRT compared the full model (with all predictors) to a reduced model (without the distance term), determining the contribution of distance to the overall model fit. Differential expression of markers was estimated based on the LRT, and significant markers were identified with adjusted p-values (FDR < 0.05). I stratified the analysis by *TREM2* [R47H>R62H>CV] variants to understand the differential impact of these variants on gene expression, separately for microglia ("Microglia C1q⁺" and "Microglia HLA-

DR⁺) (n_{CV}=6927, n_{R62H}= 3376, n_{R47H}= 1673) and astrocytes ("GFAP/S100β⁺" and "Neuro/astro") (n_{CV}=8944, n_{R62H}= 2990, n_{R47H}= 7911).

5.2.10 Density of cell types per plaque niches

I defined the density of cell type according to the following formulas:

$$\text{Density of cell type}_x \text{ in the niche} = \frac{\text{Cell count of cell type}_x}{\text{total number of cells in the niche}}$$

$$\text{Density of microglia in the niche} = \text{Density}_{\text{Microglia HLA-DR}^+} + \text{Density}_{\text{Microglia C1q}^+}$$

$$\text{Density of astrocytes in the niche} = \text{Density}_{\text{Astrocyte GFAP/S100B}^+} + \text{Density}_{\text{Astro/Neuro}}$$

5.2.11 Pseudotime of plaque niches

I computed niche-level summary statistics, including mean levels of GFAP, C1q, ApoE, S100β, AT8, HLA-DR, ALDH1L1, CD68, APP, Iba1, Sigma1R, size of cells in the niche, count of cells in the niche and the plaque area. Niches were filtered to those with a non-zero density of microglia and astrocytes (1123/4035) and categorised as amyloid-only [non-neuritic] or neuritic. I used the *umap()* function from the *umap* package (v0.2.10.0)³⁰⁸ to reduce the dimensionality of plaque-level data.

The approach uses the same concept as in pseudotime inference with scRNA-seq¹⁶⁹, except here, each data point represents a plaque niche, and pseudotime represents an inferred trajectory along which the plaque niche progresses from early to advanced stages. The plaque niche's trajectory was derived using the slingshot algorithm¹³⁵ on a two-dimensional UMAP of plaque niches. I clustered the latent space (UMAP1/UMAP2) into 3 groups via k-means clustering and chose the root of the slingshot trajectory as the cluster with a high density of amyloid niches (cluster 1) and ending cluster enriched for neuritic niches and R47H variants (cluster 3) (**Supplementary figure 5-3**). This was specified in the slingshot parameters (start.clus=1, end.clus=3).

Representative images of plaque niches [A170.13_MTG_001_8, A013.03_MTG_005_8, 20130325_MTG_002_7, 19910006_SSC_004_26] at various pseudotime values were visualised using the *PlotPixels()* from the *Cytomapper* (v1.6.0)³⁰⁹ package.

5.2.12 Cell-cell interaction

A neighbourhood interaction graph was generated via the Steinbock workflow via the “*steinbock measure neighbors --type centroids --dmax 15*” command. Cells are considered neighbours if their centroids are within a 15 μm radius of each other based on Euclidean distance, which is the default Steinbock setting.

To count pairwise interactions among all cell types in the IMC data, the *testInteractions()* function was applied (*group_by* = niche, *method* = "classic"). Per the grouping level, the plaque niche is specified here, and the function retrieves the pairwise interactions of every cell within the niche with its neighbouring cells. It computes an averaged *celltype_A-celltype_B* interaction count that can be interpreted as the average number of *celltype_B* neighbours with which a *celltype_A* cell interacts within the niche. To assess significance, interaction counts are tested against an empirical null distribution generated by permuting all cell labels, creating spatial randomness, a method established by Schapiro et al.³¹⁰. The observed interaction count is compared against the null to define significant interactions/avoidance or no significance. Representative images of plaque interactions were visualised using the *PlotPixels()* from the *Cytomapper* package (v1.6.0)³⁰⁹ colouring nodes by cell type and specifying *colPairName* = “neighborhood”.

5.2.13 Cell densities and cell-interactions along pseudotime

I used a series of linear mixed models with a random intercept to assess changes in the microglia-to-astrocyte ratio, glia densities, and cell-cell interactions in the niche as a function of pseudotime. Sample ID (image) was included as a random intercept, and *TREM2* variant and brain region were included as fixed predictors covariates. The following dependent variables were tested:

- Microglia-to-astrocyte ratio (n=1123 plaques).
- Neurotoxic and neuroprotective densities: The analysis focuses on the balance of these two subtypes in plaques containing both protective and neurotoxic astrocytes (n=409 plaques).
- Pairwise averaged cell-cell interactions in the niche among microglia/astrocyte/neuronal cell types. Sensitivity analyses included changing the independent variable from pseudotime to 4G8 and AT8.

5.2.14 Code availability

The code for this chapter can be found in the Chapter 5 folder of the Github repository <https://github.com/eleonore-schneeg/PhD>.

5.3 Results

5.3.1 IMC microglia and astrocyte subpopulations recapitulate snRNA states

272 IMC images were obtained using samples acquired from 43 post-mortem AD brains. Samples were collected from two brain regions, MTG and SOM, plus 3-4 regions of interest (ROI). Samples were stained with a panel of 24 metal-conjugated antibodies and images using the Hyperion system. IMC images were pre-processed using the Steinbock workflow for cell segmentation and intensity quantification. Amyloid plaques were segmented via Fiji, and amyloid plaques were labelled as neuritic ($4G8^+/AT8^+$) or amyloid-only [non-neuritic] ($4G8^+/AT8^-$). Masks were expanded at various distances around the core plaque 10, 20, 30, 40, 50 μm .

Using the Seurat package, I clustered 79,795 single cells with intensity measurements from 24 selected protein markers. At a 0.1 resolution, I found 8 clusters assigned to major brain cell types, including astrocytes, oligodendrocytes, neurons, and microglia, using a marker expression heatmap (**Figure 5-2 a,b**). I found two microglia subpopulations labelled “Microglia C1q⁺” and “Microglia HLA-DR⁺”, both showing co-expression with pan-microglial markers Iba1 (**Figure 5-2 c**). C1q⁺ microglia also expressed the LAMP1 lysosome

marker, potentially indicating microglia engulfing synapse. I also found two astrocyte subpopulations labelled as “Astrocyte GFAP/S100 β ”, characterised by the expression of common reactive astrocyte markers GFAP and S100 β , with the latter able to capture astrocyte processes³¹¹. The second identified astrocyte population showed co-expression with the neuronal marker MAP2 and was hence labelled as the “Neuro/Astro” population, with high expression of all astrocyte protein markers GFAP, ALDH1L1 and S100 β , as well as APOE and the immune marker IP10 (**Figure 5-2 d**). Other clusters comprised oligodendrocytes (Olig2⁺ and PLP⁺), tau expression neurons (MAP2⁺/AT8⁺) and endothelial cells (GLUT1⁺). **Table 5-3** displays these identified populations, their observed proportions, and the expected proportions from the literature. Overall, I found that the identified cell populations are consistent with expected proportions (**Table 5-3**, see “expected proportion” column), except for the “Neuro/Astro” cluster, which contained cells expressing markers that pertain to both astrocytes (ALDH1L1, APOE, GFAP, S100 β) and neurons (NeuN, MAP2). This study considers these cells to be astrocytes located near neurons; however, the IMC resolution does not allow us to differentiate them into separate populations.

I sought to validate microglia and astrocyte subpopulations defined from snRNA-seq in **Chapter 3**, section 3.3.4 (**Figure 3-5 d-f**), but here the spatial proteomics level using IMC. I first checked the expression of selected IMC markers in microglia and astrocyte snRNA-seq data from the same donors, which indicated subtype specificity only for the Micro2 subtype (**Figure 5-2 e-f**). To further quantify the concordance of IMC and snRNA-seq glia subclusters, I selected participants from the internal cohort with both IMC and snRNA-seq measurements. I computed the combined expression scores of selected markers in the two modalities and tested the difference between the snRNA and IMC score for each snRNA microglia (**Figure 5-2 g**) and astrocyte (**Figure 5-2 h**) subclusters. I expected no significant score difference for an IMC cell population that matches a snRNA subcluster. Micro2, previously characterised as an activated microglia subpopulation in **Chapter 3** corresponded to two separate IMC microglia clusters, Microglia C1q⁺ and Microglia HLA-DR⁺, which could indicate spatial differences for these activated microglia (**Figure 5-2 g**). For astrocytes, the neuro/Astro IMC cluster (ALDH1L1⁺/APOE⁺/MAP2⁺) corresponded with mixed intermediate reactive/exhausted astrocyte subpopulations (Astro3&6), while the Astrocyte GFAP/S100 β ⁺ IMC cluster matched the classical reactive astrocyte subpopulations R1/R2 (Astro2&5) (**Figure 5-2 h**). In **Chapter**

3, Astro 3&6 (mixed reactive/exhausted) showed higher expression of neurotoxic genes, while Astro2&5 (classical reactive) exhibited a reduction in neurotoxicity with an overall shift towards neuroprotection.

IMC cluster	Main marker(s) expressed	Cell type	Observed proportion	Expected proportion
Microglia (C1q ⁺)	C1Q, AT8	Microglia	0.03	12-20% ³¹²
Microglia (HLA-DR ⁺)	HLA-DR, Iba1, CD68, CSF1R		0.12	
Astrocyte (GFAP/S100β ⁺)	GFAP, S100β	Astrocyte	0.09	15-33% ³¹²
Neuro/Astro	ALDH1L1, APOE, MAP2	Mixed population: Astrocytes close to neurons	0.25	
Neuro/Tau	AT8, PHF1, MAP2	Neuron	0.26	
Oligo (Olig2 ⁺)	OLIG2	Oligodendrocyte	0.06	12-42% ³¹²
Oligo (PLP ⁺)	PLP		0.12	
Endothelial (GLUT1 ⁺)	GLUT1	Endothelial	0.07	0.2-7% ^{313,314}

Table 5-3 Observed and expected cell proportions

Expected proportions (except endothelial cells) were estimated from cell deconvolution in the human adult prefrontal cortex³¹².

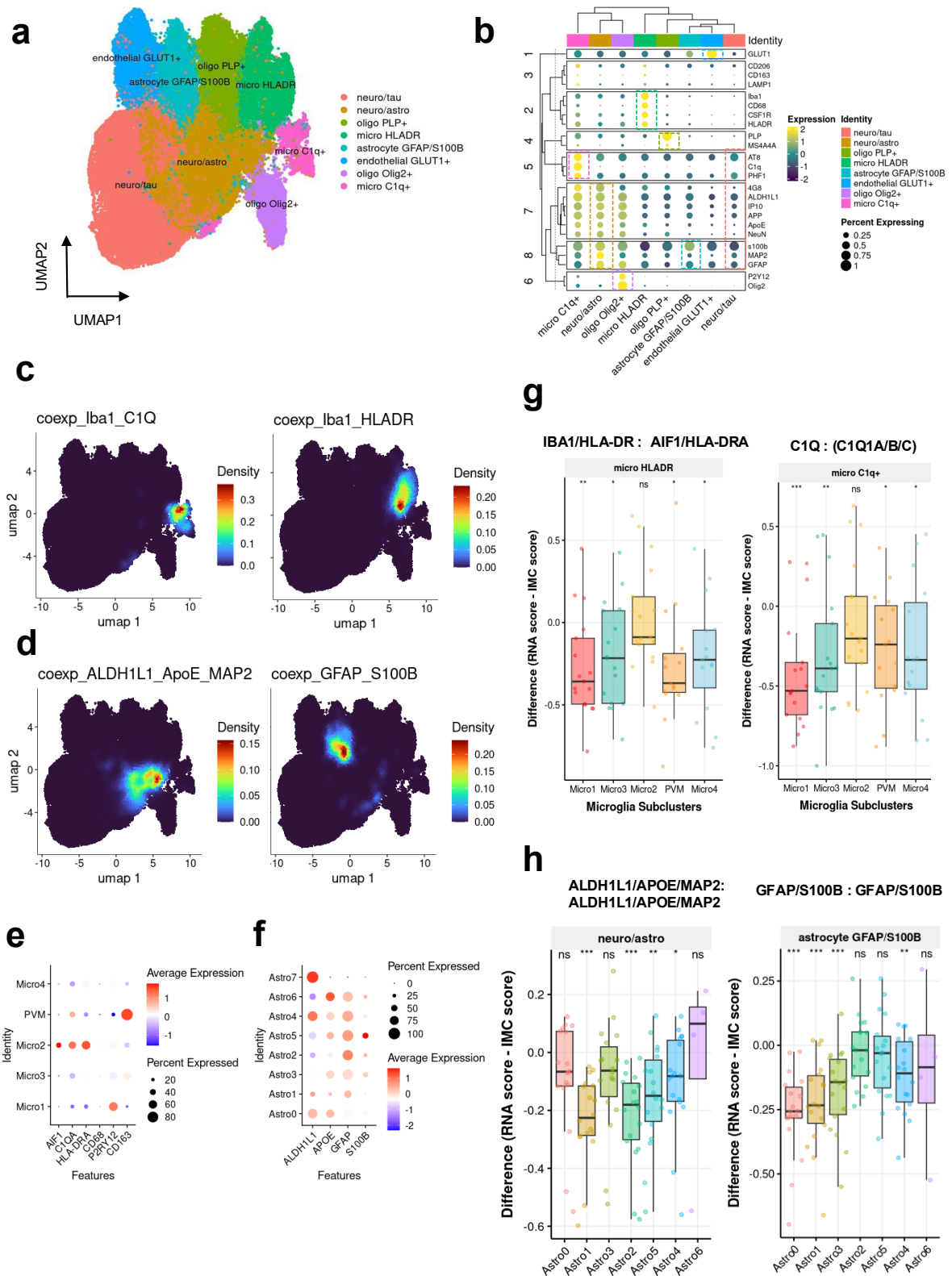


Figure 5-2 Spatial omic characterisation of glial activation in AD using imaging mass cytometry

a) UMAP plot displaying 8 clusters assigned to major brain cell types identified from single cells segmented from IMC images. **b)** Heatmap illustrating the expression of key markers across the 8 cell clusters. **c)** Co-expression densities of Iba1 and C1q (left) and Iba1 and HLA-

DR (right). **d)** Co-expression densities of *ALDH1L1*, *APOE* and *MAP2* (left) and *GFAP* and *S100 β* (right). **e)** IMC microglia markers expression in snRNA-seq clusters studied in Chapter 3. UMAP projections display the expression levels of these microglial markers *Iba1*, *HLA-DR*, and *C1q* in IMC. **f)** Same as in panel **e)** but for astrocyte markers *ALDH1L1*, *APOE*, *S100 β* and *GFAP*. **g)** Box plots comparing the expression scores derived from IMC and snRNA-seq microglia subclusters. Scores are computed for Microglia *HLA-DR*⁺ (left) and Microglia *C1q*⁺ (right). **h)** Same as in panel **g)** but for astrocytes subclusters *Neuro/Astro* (left) and *GFAP/S100 β* ⁺ (right).

5.3.2 Molecular and cellular footprinting of plaque niches in AD

As a validation for plaque segmentation, I compared the current AD plaque data to control data. I found a significant increase in all amyloid plaque in AD vs. controls (**Figure 5-3 a**). The proportion of neuritic plaques out of all plaques was significantly higher in AD versus controls, stratified by *TREM2* variants (**Figure 5-3 b**). There was a variant-dependent relationship by severity of *TREM2* variants, where the R47H has around 50% neuritic plaques compared to 25% in the common variant (CV) (**Figure 5-3 c**).

Next, I investigated the spatial relationship between plaques and surrounding cellular populations. In **Figure 5-3 d:h**, I showed the plaque and cell segmentation workflow, which allowed to tag the cells on plaques. Cells were localised to the closest plaque and assigned a distance. A plaque niche was defined as the core area expanding by 10 μ m. Previous spatial analyses of AD have used varying definitions of the plaque niche, usually between 10 and 50 μ m. I defined the distance threshold for a plaque niche to 10 μ m, given that this region showed the most variation in the relative proportion of cells, indicating a plaque influence on the cell states (**Figure 5-3 e**). Using this definition, I identified 4035 distinct plaque niches in the images: 2653 amyloid-only [non-neuritic] niches (4G8⁺/AT8⁻) and 1382 neuritic niches (4G8⁺/AT8⁺). Altogether, 17,415 cells were located within these niches across all images, including 7599 cells in neuritic niches and 9816 cells in amyloid-only [non-neuritic] niches. Finally, I found that the glial cells Microglia *C1q*⁺, Microglia *HLA-DR*⁺ and *Neuro/Astro* were predominantly found within 10 μ m of plaques. Their relative proportion gradually decreased as the distance to the plaque increased. The trend was consistent between amyloid-only and neuritic plaques (**Figure 5-3 i**). The other populations identified had a reverse trend whereby their relative proportions tended to be the lowest close to plaques and increased with plaque distance.

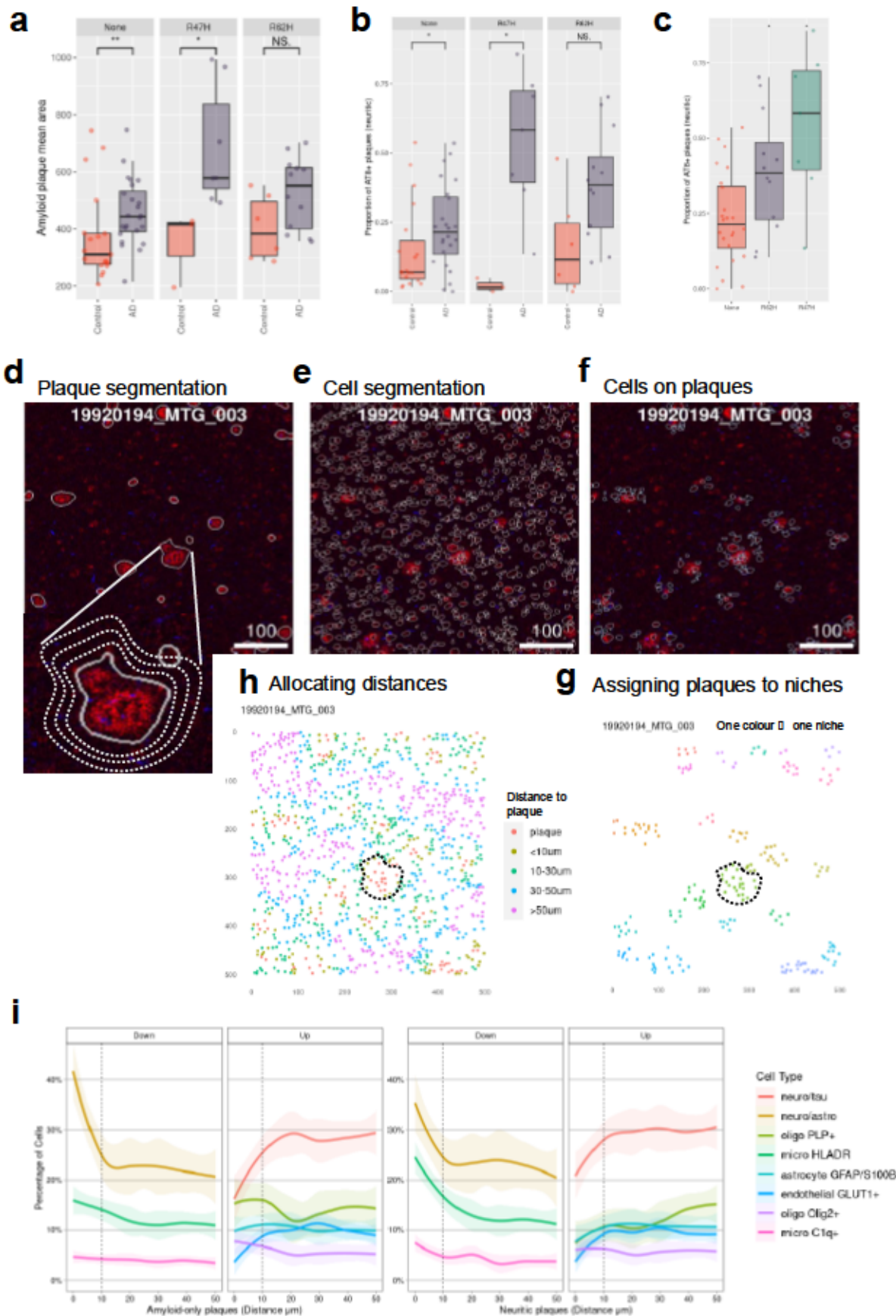


Figure 5-3 IMC allows spatial characterisation of amyloid plaques and cells

a) Boxplot showing the mean area of amyloid plaques (all plaques) in AD cases versus controls, stratified by different *TREM2* variants. **b)** Boxplot showing the proportion of neuritic plaques (4G8+/AT8+ plaques) out of all plaques, compared between AD and controls. **c)** relationship between *TREM2* variants and the proportion of neuritic plaques. P-values derived from t-test relative to CV (* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$). **d)** Schematic of the plaque and cell segmentation workflow. Plaques were segmented using 4G8 (amyloid) and AT8 (neuritic) markers, and **e)** cells were segmented, **f)** located on plaques and **g)** assigned to plaque niches. **h)** Each cell was assigned to a distance to the nearest plaque. The plaque niche was defined as the core area expanding by 10 μm . IMC images revealed distinct plaque niches across 4035 regions (2653 amyloid-only and 1382 neuritic plaques), containing 17,415 cells. **i)** Loess-fitted curves showing the relative proportions of identified cell types (e.g., Microglia C1q+, Microglia HLA-DR+, Neuro/Astro) within increasing distances from all amyloid-only (left) and neuritic plaques (right). Cells such as Microglia and Neuro/Astro populations show higher proportions within 10 μm of plaques (“Down”). In contrast, other cell types show an inverse relationship, with their relative abundance increasing as the distance from plaques increases (“Up”).

5.3.3 Glial activation states show spatial variability relative to plaque niches

Next, I wanted to characterise this spatial variability in more detail and according to both the plaque phenotype and *TREM2* risk variants. I conducted permutation test analyses to compare the proportions of different cell types within the plaque niche versus non-niche cells across different plaque phenotypes (amyloid-only [non-neuritic], neuritic, all plaques) (**Figure 5-4 a**). The observed Log₂ fold difference (obs_log₂FD) represents the Log₂ ratio of the proportion of a specific cell type within the plaque niche compared to its proportion outside the niche, indicating the enrichment of that cell type within plaque niches relative to areas outside the niche. Generally, the glial cells Microglia C1q⁺, Microglia HLA-DR⁺ and Neuro/Astro had an increased proportion in the niche compared to outside of the niche, and the log₂FD was higher in neuritic niches than amyloid-only niches. For instance, Microglia HLA-DR⁺ in the neuritic niches had a Log₂FC of close to 1, representing twice the proportion found outside the niche. When comparing *TREM2* variants relative to CV, both had a decreased relative proportion of Microglia C1q⁺ within the neuritic plaque niche. R47H variants had a reduced proportion of Microglia HLA-DR⁺ and an increased proportion of Neuro/Astro relative to CV. The opposite trend was seen in R62H variants (**Figure 5-4 b**). Outside the niche, I found more Tau⁺ neurons in both variants and fewer classical reactive GFAP/S100 β ⁺ astrocytes (**Figure 5-4 c**).

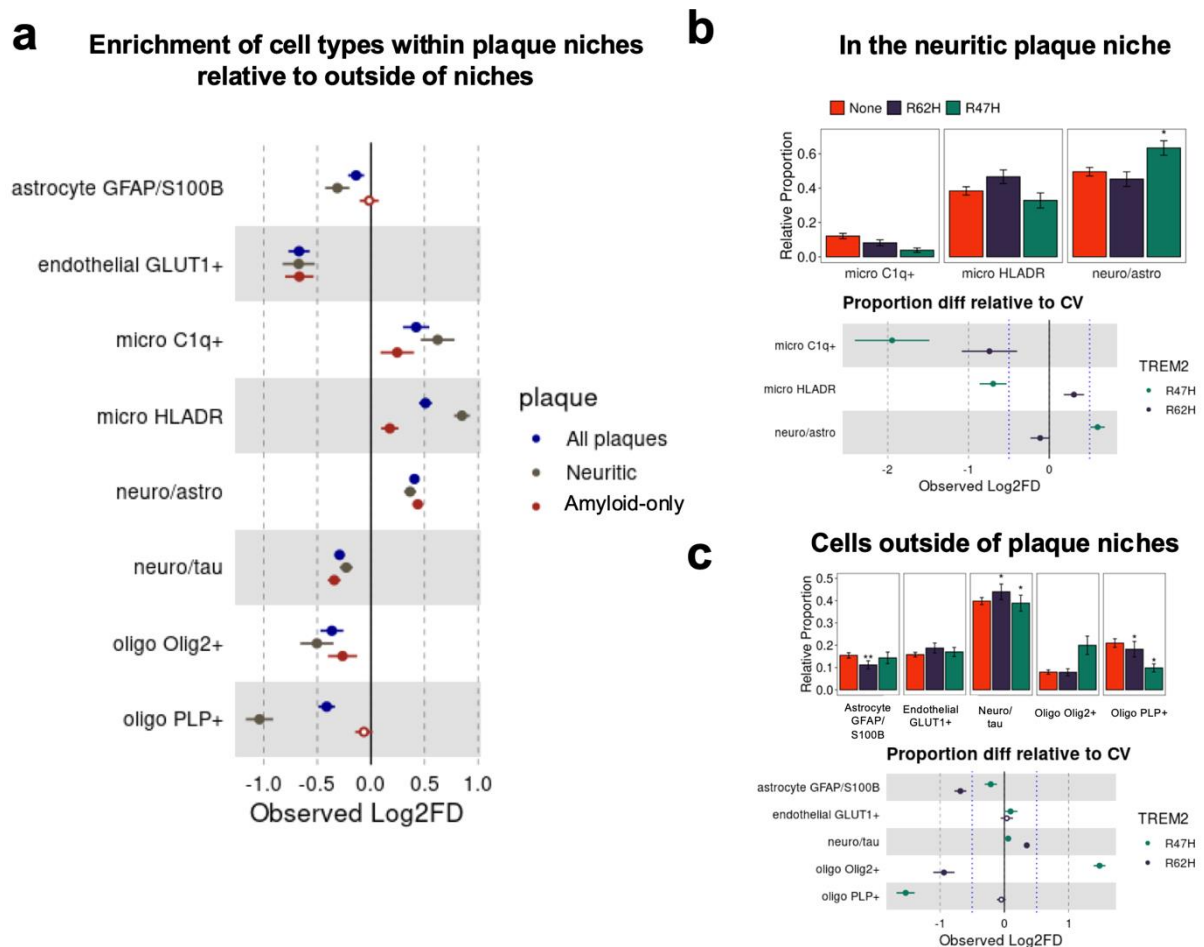


Figure 5-4 Microglia and Astrocyte are predominantly found in the niche

a) Permutation tests comparing the enrichment of specific cell types within amyloid plaque niches versus outside of the plaque niches across all plaque phenotypes (amyloid-only, neuritic, and all plaques). Analysis was performed on all cells ($N=79,795$) and subsets removing either neuritic ($N=72,196$) or amyloid-only niches ($N=69,979$). Positive $\log_2FD$ values indicate enrichment within the plaque niche, while negative values indicate depletion.

b) Upper panel: Bar plots showing the relative proportions of cell types predominantly enriched in neuritic plaque niches. The lower panel shows a permutation test comparing the proportions of these cell types in *TREM2* risk variants (R47H and R62H) relative to CV. The total number of cells analysed for this comparison is $n=4099$ (i.e. in niches and for selected cell subtypes).

c) Same as in b) but looking at cells found outside of niches. The total number of cells analysed for this comparison is $n=39,289$ (i.e. outside niches and for selected cell subtypes).

Next, I wanted to characterise this spatial variability in protein expression from microglia (Microglia $C1q^+$, Microglia $HLA-DR^+$) and astrocytes (Neuro/Astro, Astrocyte $GFAP/S100\beta^+$), stratified by *TREM2* variants. I used linear mixed models for each *TREM2* variant with a random intercept ($1|sample_id$) to model the single-cell protein marker expression change relative to the increasing distance to plaque. I found a consistent up-regulation of microglia markers in cells approaching plaques in all *TREM2* variants, including

CD68 and HLA-DR. APOE was up-regulated in *TREM2*^{R62H} microglia. CSF1R was found to be up-regulated only in *TREM2* variants. For astrocytes, GFAP was up-regulated in *TREM2*^{R47H} but down-regulated in CV. S100 β was down-regulated when cells are approaching plaques, thereby reiterating the finding that GFAP/S100 β ⁺ astrocytes are predominantly found outside of plaque niches.

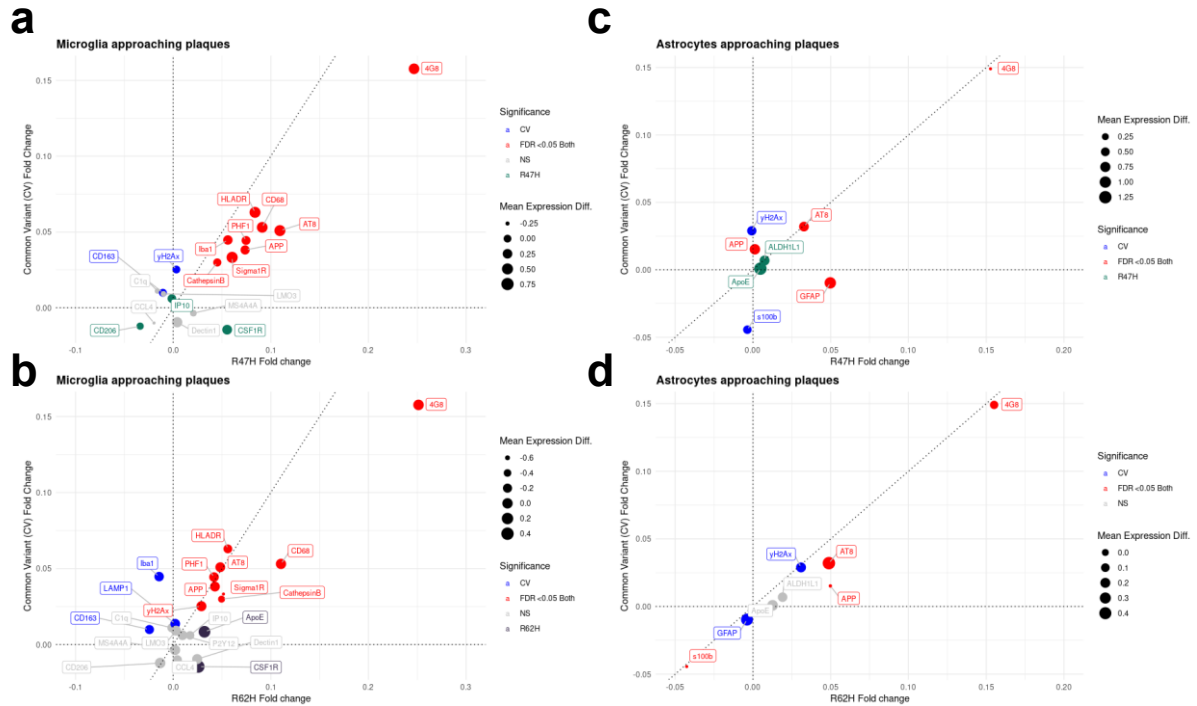


Figure 5-5 Spatial variability in protein expression

a-b) Differential expression analysis of IMC markers using a generalised linear mixed model for microglia ($n=11,976$) (a-b) and astrocytes ($n=19,845$) (c-d) relative to their distance from plaques. The model assesses the fold change in marker expression as a function of proximity to plaques, with analyses stratified by *TREM2* variants. The upper panel plots show fold changes for the R47H variant compared to CV, while the bottom panel illustrates fold changes for the R62H variant compared to CV. Colour represents significance ($FDR < 0.05$) in CV, *TREM2* variants, both, or N.S in grey.

5.3.4 Imbalance in astrocyte states and altered cell-cell interactions within the plaque microenvironment along the plaque trajectory

I then investigated the evolution of cell-cell interaction within the plaque microenvironment, to study how glial subpopulations crosstalk dynamically impact AD pathology progression. I derived a UMAP of unique plaque niches with the criteria to have non-zero microglia and astrocyte density ($n=1123/4035$). The summed cell size, cell count, and normalised plaque area,

neuritic status, were considered along the average expression of GFAP, C1q, ApoE, S100 β , 4G8, AT8, HLA-DR, ALDH1L1, CD68, APP, Iba1, and Sigma1R for the UMAP dimensionality reduction. Plaque pseudotime was significantly associated with plaque size (p-value <0.001), highlighting that plaque trajectory can be a good proxy for studying its evolution or severity. In general, plaques from *TREM2* samples and neuritic plaques, which depict more advanced stages of an AD plaque, were found at higher pseudotime values (**Figure 5-6 a-b**). Pseudotime was regressed against the microglia-to-astrocyte density ratio, and the significant positive relationship suggests that, as plaque pseudotime increases, niches become denser in activated microglia relative to astrocytes. Examining the dynamics within the astrocyte subpopulation in greater detail, I found that there was a significant decline in the density of classical reactive astrocytes (**Table 5-4**).

I then wanted to study the evolution of cell-cell interactions along pseudotime. I quantified cell-cell interactions within individual niches whereby cells were interacting neighbours if their centroids were within 15 μ m of each other (**Figure 5-6 c**). I derived a celltype_A-celltype_B interaction score for each type of interest, representing the number of neighbours of type “celltype_B” that celltype_A has in each niche (**Figure 5-6 d**). As I explored cell-cell interactions along plaque pseudotime, I observed significant changes in the levels of these interactions. Specifically, I identified a decrease in C1q⁺ microglia to GFAP/S100 β ⁺ astrocyte interactions, while there was an increase in interactions between microglia C1q⁺ and neurotoxic astrocytes (Neuro/Astro). This suggests that, as plaques progress and pseudotime advances, there is a shift from overall protective astrocyte-microglia interactions towards more harmful microglia-astrocyte interactions. Examining the morphology of neurotoxic mixed reactive/exhausted astrocytes, all classified as Neuro/astro in the donor A013.03_MTG_005 of **Figure 5-6 c**, I found that these closely resembled the morphology of a distinct subset of astrocytes characterised in Sokolova et al. 2024, which are distinguished by their “deformed” and “spherical” peri-synaptic processes, or so-called “bulbous” morphology⁴⁵ (**Figure 5-7**). As a sensitivity analysis, cell-cell interactions were regressed against the mean level of 4G8 in the plaque, which yielded negative associations with all cell-cell interactions with Tau⁺ neurons or GFAP/S100 β ⁺ reactive astrocytes. For instance, neuronal interactions, such as neuro/tau to neuro/tau, decreased as amyloid [4G8] increased, which may indicate neuronal loss as plaques develop. Similarly, GFAP/S100 β ⁺ reactive astrocytes tended to interact less with Tau⁺ neurons.

Overall, the data imply that local plaque-related toxicity is exacerbated, in part, by the mixed reactive/exhausted astrocyte population, which contributes to changes in cell-cell interactions within the plaque microenvironment, further highlighting the crosstalk between pathology, microglia-astrocyte and glia-neuronal dynamics in AD progression.

<i>Predictors</i>	Ratio [micro density/Astro density]			Microglia density			Classical reactive astrocyte density [GFAP/ S100β ⁺]			Mixed reactive/exhausted astrocyte density [Neuro/Astro]		
	<i>Est</i>	<i>CI</i>	<i>p</i>	<i>Est</i>	<i>CI</i>	<i>p</i>	<i>Est</i>	<i>CI</i>	<i>p</i>	<i>Est</i>	<i>CI</i>	<i>p</i>
(Intercept)	0.9 7	0.79 – 1.1 6	<0.001	- 0.0 0	- 0.18 – 0.1 7	0.96 0	0.7 5	0.66 – 0.8 4	<0.001	0.1 3	0.08 – 0.1 8	<0.001
pseudotime	0.0 2	0.01 – 0.0 3	0.007	0.0 1	- 0.00 – 0.0 3	0.050	- 0.0 1	-0.01 – - 0.00	0.015	0.0 0	- 0.00 – 0.0 1	0.294
<i>TREM2</i> [R62H]	0.2 9	0.05 – 0.5 2	0.017	0.0 5	- 0.17 – 0.2 7	0.65 3	- 0.0 2	- 0.14 – 0.1 1	0.773	0.0 1	- 0.07 – 0.0 8	0.844
<i>TREM2</i> [R47H]	- 0.3 4	-0.59 – - 0.09	0.009	- 0.2 6	-0.49 – - 0.02	0.034	- 0.1 5	-0.28 – - 0.01	0.038	0.1 1	0.03 – 0.2 0	0.007
Brain Region [SOM]	- 0.0 1	- 0.21 – 0.1 9	0.929	- 0.1 7	- 0.35 – 0.0 2	0.07 9	- 0.0 4	- 0.14 – 0.0 5	0.374	0.0 3	- 0.03 – 0.0 8	0.362
Random Effects												
σ ²	1.00			0.84			0.07			0.02		
τ ₀₀	0.17 sample id			0.16 sample id			0.03 sample id			0.01 sample id		
ICC	0.14			0.16			0.29			0.36		
N	209 sample id			209 sample id			114 sample id			114 sample id		
Observations (plaques)	1123			1123			409			409		
Marginal R ² / Conditional R ²	0.038 / 0.178			0.023 / 0.180			0.062 / 0.338			0.065 / 0.403		

Table 5-4 Changes in niches' microglia/astrocyte density ratio

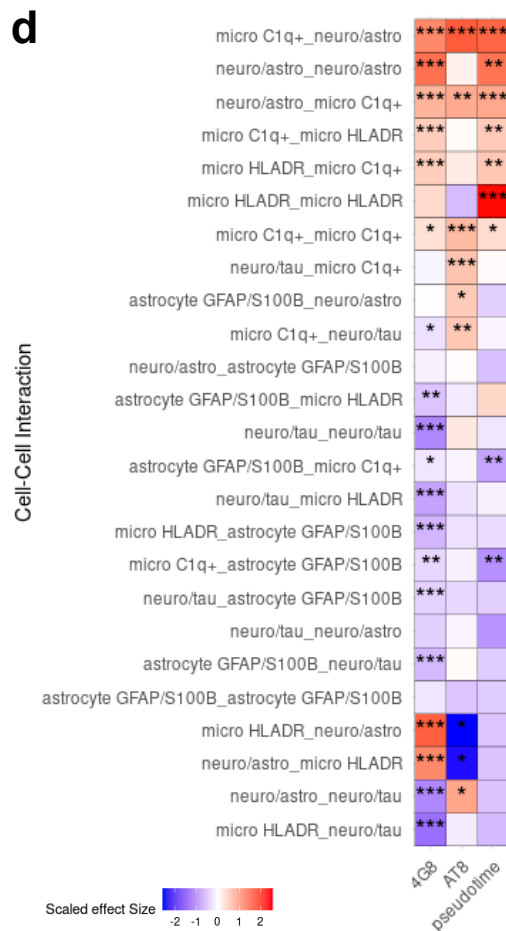
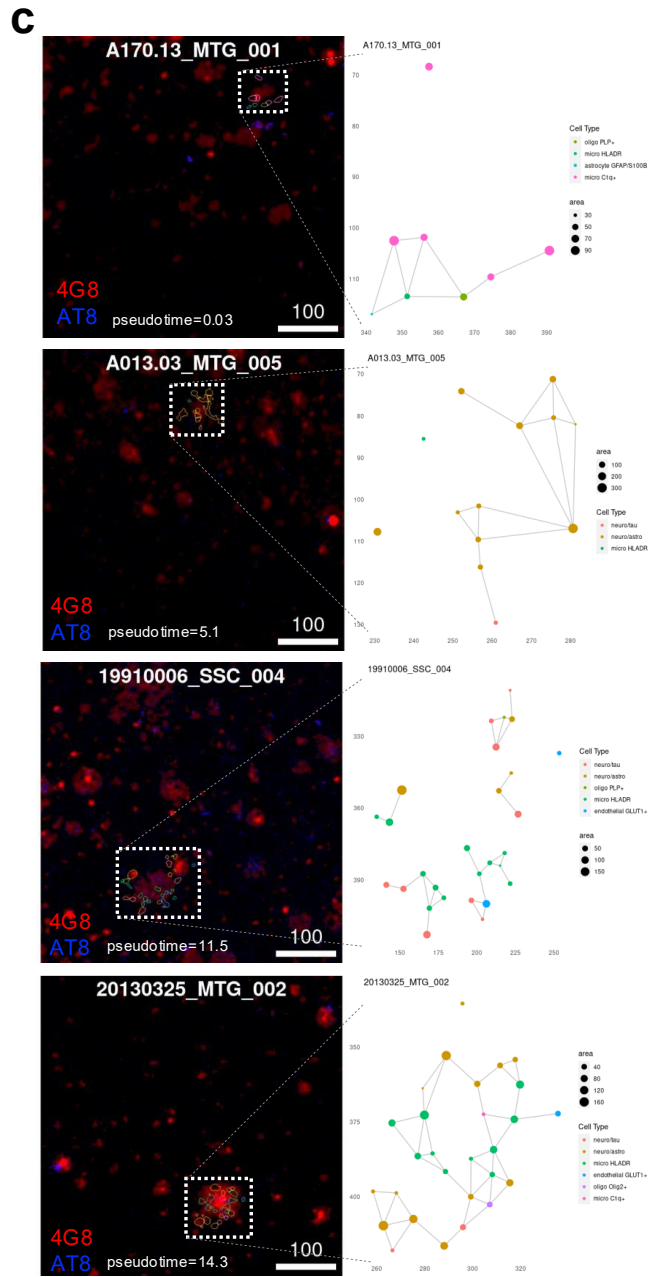
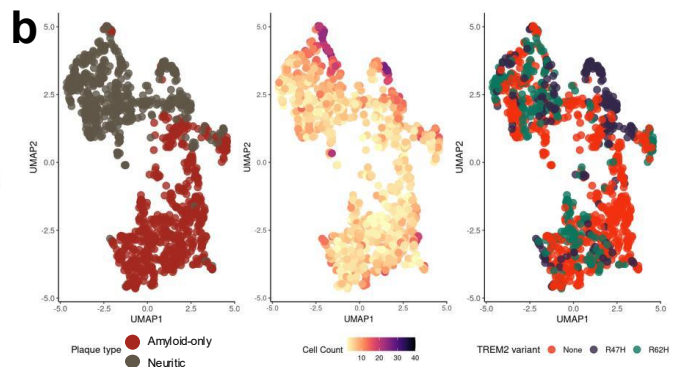
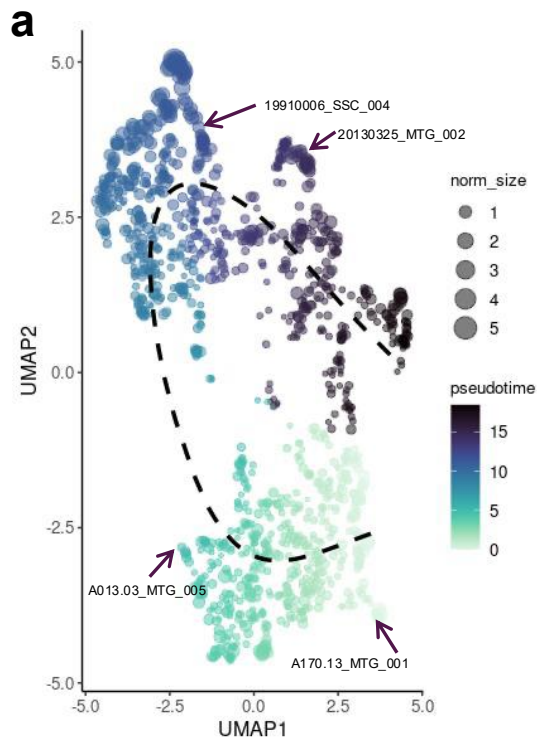


Figure 5-6 Evolution of cell-cell interactions within the plaque niche along pseudotime
a) UMAP of plaque niches, each dot corresponds to a plaque niche from IMC images. The size of the dot represents the normalised plaque area. The colour represents the pseudotime value, and the dotted line is the slingshot plaque trajectory. Arrows point to representative plaques shown in panel c). **b)** Left panel labels niches as either neuritic ($AT8^+$) or amyloid-only ($4G8^+$, $AT8^-$). The middle panel shows the number of cells present in each niche. The right panel shows plaques coloured by the TREM2 variant. **c)** Representative images of four plaques at pseudotime=0.03, 5.1, 11.5, and 14.3. Images are tagged with 4G8 antibody (red) and AT8 (blue). Scale bar = 100 μ m. Segmented cells appear in the selected plaque niche (white dotted box). The right panel corresponds to the modelled cell-cell interactions within the plaque niche, where cells are coloured based on cell type and sized based on cell area. **d)** Summary of linear mixed models assessing the association between cell-cell interactions and 4G8, AT8 and pseudotime. The models include sample ID as a random intercept and controls for TREM2 variant and brain region as fixed predictors. The colour scale shows the scaled effect size and p-values (* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$).

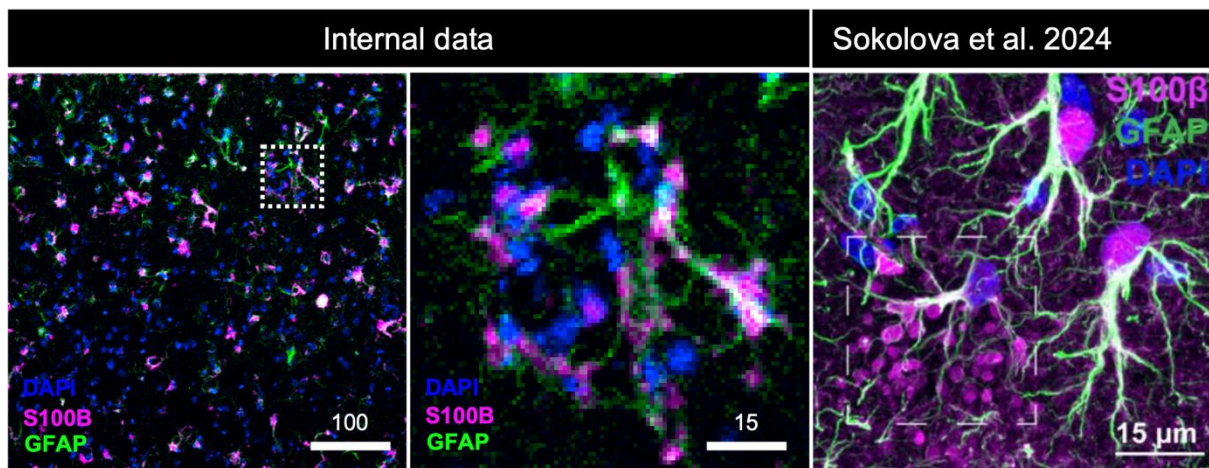


Figure 5-7 Neurotoxic astrocytes relate to the bulbous morphology characterised in an AD mouse model

IMC image taken from A013.03_MTG_005 also displayed in Figure 5-6. All astrocytes from this image were categorised as neuro/astro. Scale bar = 100 μ m. The right panel shows a high-resolution confocal microscopy image from Sokolova et al. 2024, displaying an AppNL-F knock-in mouse hippocampus labelled with S100 β in purple and GFAP in green.

5.4 Discussion

In this study, I leveraged IMC data to analyse brain tissue from two regions of AD donors, using a panel of over 40+ metal-conjugated antibodies. This enabled to identify glial cell phenotypes, subpopulations, and their spatial distributions relative to amyloid plaques at the single-cell resolution. The antibody panel included cell-specific and functional markers, facilitating the exploration of single-cell phenotypes, functional states, and cell-cell interactions within the complex tissue microenvironment.

5.4.1 Distinct glial subpopulation and spatial alterations in AD plaque microenvironment

This analysis identified distinct glial phenotypes, which I mapped to those previously identified through snRNA-seq in **Chapter 3**. Specifically, I identified two subtypes of activated microglia: "C1q⁺" and "HLA-DR⁺" microglia. I also identified two astrocyte subpopulations: (1) "GFAP/S100 β ⁺" classical reactive astrocytes, associated with an overall net neuroprotective effect, and (2) a "Neuro/Astro" astrocytes population localised near neurons, characterised by mixed intermediate reactive and exhausted features, contributing to a net neurotoxic effect. The enrichment of activated microglia and mixed reactive/exhausted astrocytes within amyloid plaque niches corroborates previous findings that suggest a localised detrimental cellular environment around plaques^{43,284}. I observed that proximity to plaques significantly influenced protein expression within these cells, underscoring the importance of spatial context in AD pathology.

Further investigating the spatial relationships within amyloid plaque niches, I observed that AD samples displayed significantly more plaques than controls, and the proportion of neuritic plaques was increased in AD patients carrying *TREM2* risk variants (R47H/R62H) compared to those with CV. This aligns with previous findings from a study conducted in the Matthew's lab using snRNA-seq data³¹. These results are consistent with the critical role of the *TREM2* receptor in microglial activation and function. It regulates microglial phagocytosis of amyloid plaques and cellular debris, which is essential for mitigating AD pathology³¹⁵. *TREM2* (loss of function) variants have been associated with reduced microglial clustering around plaques, impairing A β clearance, leading to alterations in plaque morphology and increased neurotoxicity^{316,317}. These findings indicate a reduced microglia activation in *TREM2*^{R47H} donors, as evidenced by decreased proportions of both HLA-DR⁺ and C1q⁺ microglia subtypes in the plaque niche. A similar, though less pronounced, trend was observed in *TREM2*^{R62H} carriers for the C1q⁺ microglia. The increase in neurotoxic mixed reactive/exhausted astrocytes in *TREM2*^{R47H} carriers suggests a detrimental mechanism that may result from impaired microglial function and a higher local accumulation of amyloid plaques. Fancy et al. indicated that *TREM2*^{R47H} astrocytes may have a compensatory mechanism to mitigate weakened microglial responses and amyloid phagocytosis³¹. This was paralleled with a reduced proportion of classical reactive neuroprotective astrocytes (GFAP/S100 β ⁺) in *TREM2*^{R62H/R47H}

donors. This phenomenon was recorded in a recent ischemic stroke study where *TREM2* knockout in astrocytes led to an increase in the pro-inflammatory astrocyte marker C3 but decreased S100 calcium-binding (S100A10) expression on reactive astrocytes³¹⁸. Finally, the increased tau-accumulating neurons in *TREM2* donors corroborate previous research that links *TREM2* dysfunction to tau pathology³¹⁷. Although both *TREM2* R47H and R62H are associated with altered glial responses, their effects are not identical; therefore, analysing them separately provides valuable insights into glial subtype dynamics and heterogeneity in response to distinct *TREM2* variants.

5.4.2 Pseudotemporal dynamics reveal astrocyte-microglia shifts towards neurotoxicity in plaque niches

I leveraged an approach to assess spatiotemporal changes within plaque niches using a pseudotemporal analysis. This method proved to be a valuable proxy for evaluating plaque trajectory, evolution, and severity. By assigning a pseudotime value to each plaque niche, I could model the progression of cellular and molecular changes relative to plaque development. The cell-cell interaction analysis leveraged spatial proximity (cells within a maximum of 15 μm of each other) to define interactions. The results revealed an imbalance in the local astrocyte population towards a more harmful, neurotoxic mixed reactive/exhausted state as plaques progressed. This finding aligns with the work of Mallach et al., who reported that neurotoxic astrocytes disrupt synaptic signalling and adopt a more harmful state in microglia-dense regions⁴³. This could also underlie the aberrant synaptic and neuronal loss observed in these regions²⁸⁸. I found evidence of direct increased interactions between neurotoxic mixed reactive/exhausted astrocytes and tau⁺ neurons in AT8-rich plaques, that could indicate a direct harmful effect. An indirect effect could also be at play via the astro-to-micro, micro-to-neurons interaction routes, as well as an overall detrimental “neuro-neglect” effect of exhausted astrocytes.

I propose that the harmful astrocyte effect is associated with a local imbalance, whereby the density of neuroprotective classical reactive astrocytes (GFAP/S100 β ⁺) decreases as plaques evolve, leading to reduced interactions between these astrocytes and C1q⁺ microglia. Concurrently, there is an increase in interactions between neurotoxic mixed reactive/exhausted astrocytes (Neuro/Astro) and C1q⁺ microglia. This shift from protective astrocyte-microglia

interactions to more harmful microglia-neurotoxic astrocyte interactions, aligns with observations that neurotoxic astrocytes in the Neuro/Astro population exhibited morphological similarities to the bulbous astrocytes characterised in the Sokolova et al. study⁴⁵. Bulbous astrocytes are associated with decreased astrocyte-synapse coverage and an upregulation of MFGE8, which promotes microglial synapse engulfment. Interestingly, these findings reveal that C1q⁺ microglia interacting with the Neuro/Astro subpopulation also expressed LAMP1, a lysosomal marker, consistent with increased local synapse engulfment. While Sokolova et al. did not observe an overrepresentation of bulbous astrocytes near amyloid plaques, the Neuro/Astro subpopulation, which resembles this morphology, was predominantly located within the plaque niche. This suggests a regional specialisation where neurotoxic mixed reactive/exhausted astrocytes contribute to the pathological microenvironment surrounding plaques, which may exacerbate synaptic loss and neuronal dysfunction. This also reflects a potential transition where classical reactive astrocytes lose their capacity to sustain reparative functions, eventually becoming exhausted. However, the fact that the density of the mixed reactive/exhausted astrocyte population does not increase in the niche with pseudotime likely reflects that, by late stages of the disease analysed here, this population has already reached a high level of abundance. This plateau could signify that the exhaustion process occurred earlier in the disease trajectory and that these astrocytes, once exhausted, persist without further proliferation or replacement. This stagnation, coupled with the concurrent decrease in classical reactive astrocytes, underscores the failure of astrocytic reparative mechanisms as the disease progresses.

5.4.3 Study limitations

Several study limitations must be highlighted. Most of the samples analysed were from individuals in late-stage (Braak V-VI) AD rather than mid-stage (Braak III-IV), which may limit the ability of the pseudotemporal analysis to fully capture the true trajectory of disease evolution. Moreover, IMC presents analytical challenges, particularly in accurate cell segmentation and cell type phenotyping³¹⁹. Automated workflows like the Steinbock aim to streamline segmentation. However, pre-trained deep learning segmentation algorithms such as DeepCell used in this study are trained on datasets not specific to brain nor *post mortem* tissue, thereby may not fully capture brain cell morphology and staining patterns. Additionally, this

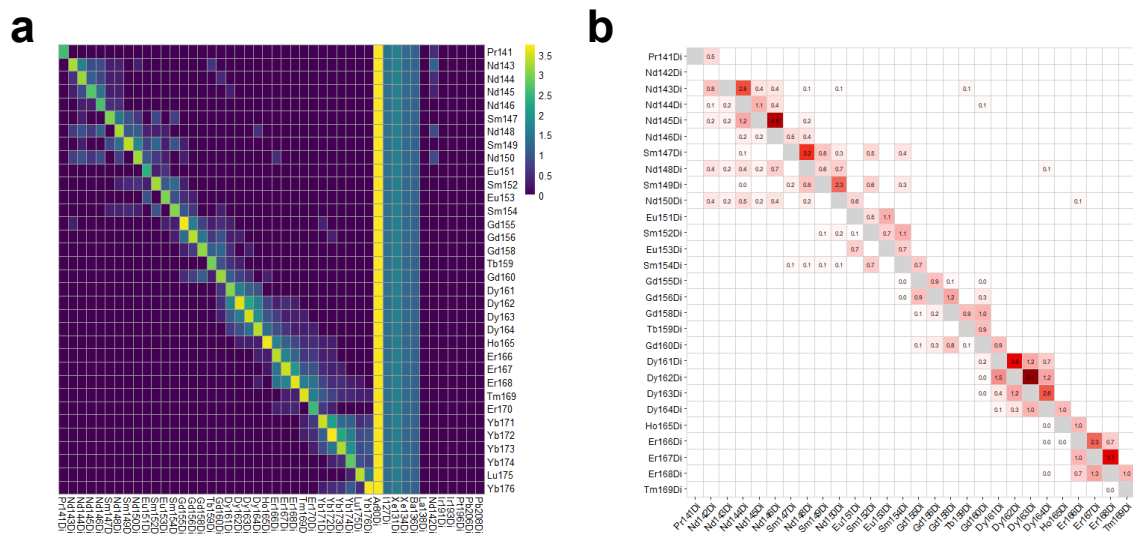
analysis relied solely on nuclear DNA staining for cell segmentation without a universal pan-cellular marker, which may lead to incomplete cell detection. Regarding cell type phenotyping, IMC's somewhat low resolution ($1\ \mu\text{m}^2/\text{pixel}$) makes differentiating cells in densely packed areas challenging. This led to the identification of a mixed population, the "Neuro/Astro" cluster, which was assumed to be an astrocyte population close to neurons based on prior knowledge of the tripartite synapse and expected cell type proportions. One should also highlight that the 2D sections lead to an incomplete picture of the 3D plaque micro-environment where signals from cellular processes extending in three dimensions, such as dendrites or astrocytic projections, might be detected within the area of neighbouring cells in the 2D plane, potentially leading to misattribution of signals and affecting the accuracy of cell type identification. This also applies to the detection of the closest plaque niche, where the assigned "closest" plaque could in fact be more distant than another closer one in the 3D space but that is not captured in the 2D slice. Spillover effects between metal-conjugated antibodies can further compromise the accuracy of protein quantification despite compensation. Finally, the use of spatial proximity within a $15\ \mu\text{m}$ threshold as a proxy for interaction may not accurately reflect functional interactions, as cells nearby may not interact, and meaningful interactions beyond this range might be overlooked.

5.5 Conclusions and future work

In conclusion, these findings validate the utility of IMC for the spatial identification of cell subpopulations in post-mortem AD brain tissues. I identified distinct activated microglia and astrocyte subpopulations and demonstrated how *TREM2* risk variants and plaque evolution influence their spatial distribution and interactions within amyloid plaque niches. This spatial remodelling of glial activation states provides insight into the dynamic nature of plaque evolution, suggesting that cellular responses within the niche are not static but adapt in relation to genetic background and plaque maturation. These findings highlight an increased microglial activation and a shift towards neurotoxic astrocyte states due to decreased neuroprotective classical reactive astrocytes in the plaque niches. Future studies should validate the present findings with higher-resolution imaging like confocal microscopy. This would provide more detailed insights into glia-glia and glia-neuronal interactions and offer an optimised view of the 3D plaque micro-environment. Moreover, while this study examined spatio-

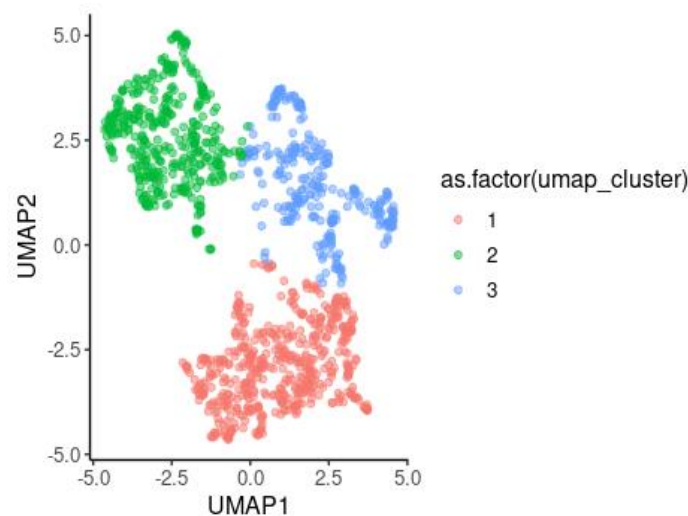
pseudotemporal relationships cross-sectionally, the inclusion of mid-stage samples would provide a more comprehensive view of disease progression. Future work could employ longitudinal studies to track plaque evolution and cellular responses over time in animal models or organoid systems.

5.6 Supplementary figures



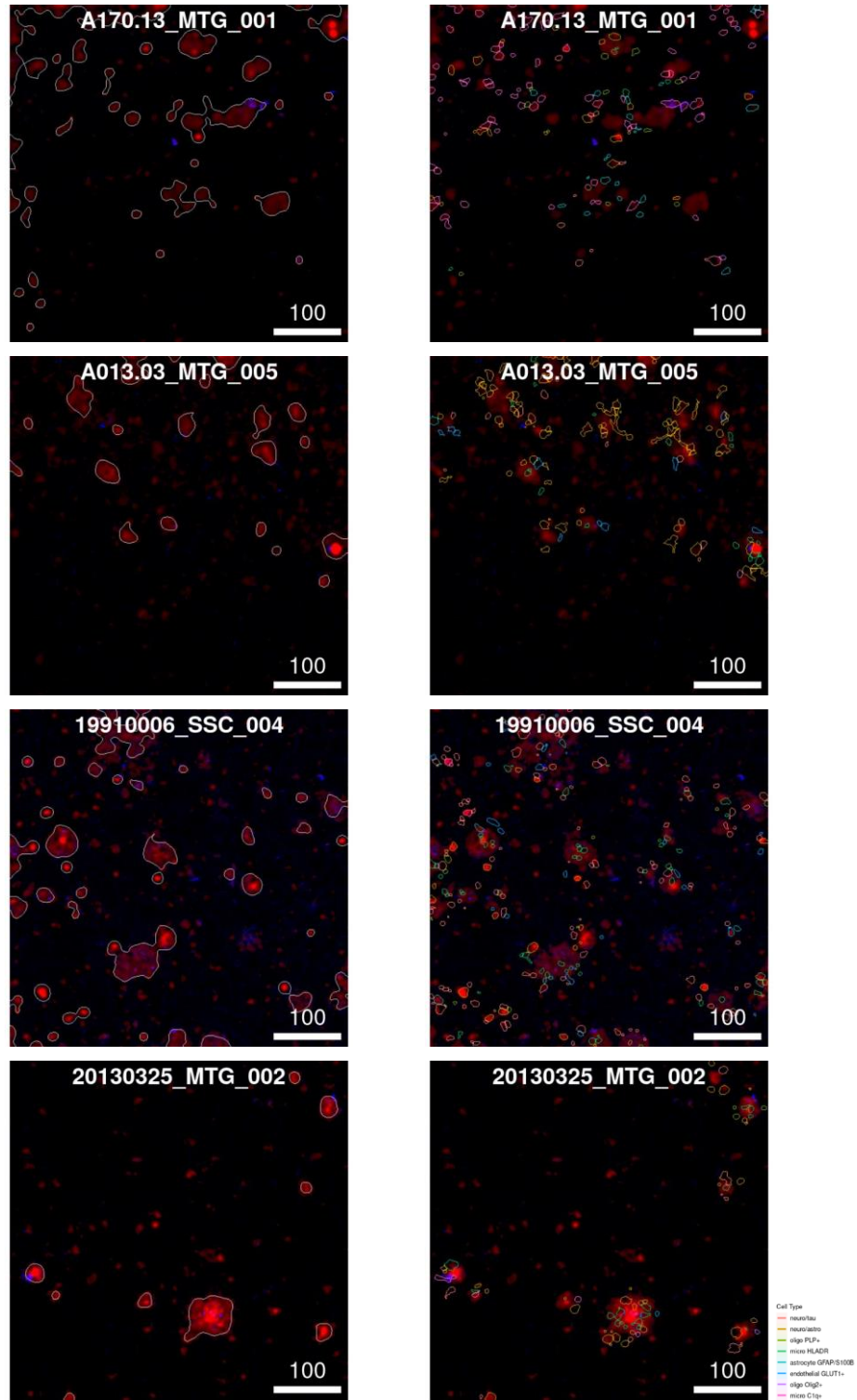
Supplementary figure 5-1 IMC spillover correction

a) Log₁₀ median pixel counts per spot and channel. **b)** Spillover matrix computed based on the single-positive pixels. Spillover is represented as the signal percentage overlap between each conjugated metal isotopes.



Supplementary figure 5-2 Niche level clustering

K-means clustering of plaque niches. The starting cluster for the slingshot trajectory is 1, and the ending cluster for the slingshot trajectory is 3.



Supplementary figure 5-3 Representative images of plaques and cells

Chapter 6 Concluding remarks and future work

Graphical Abstract

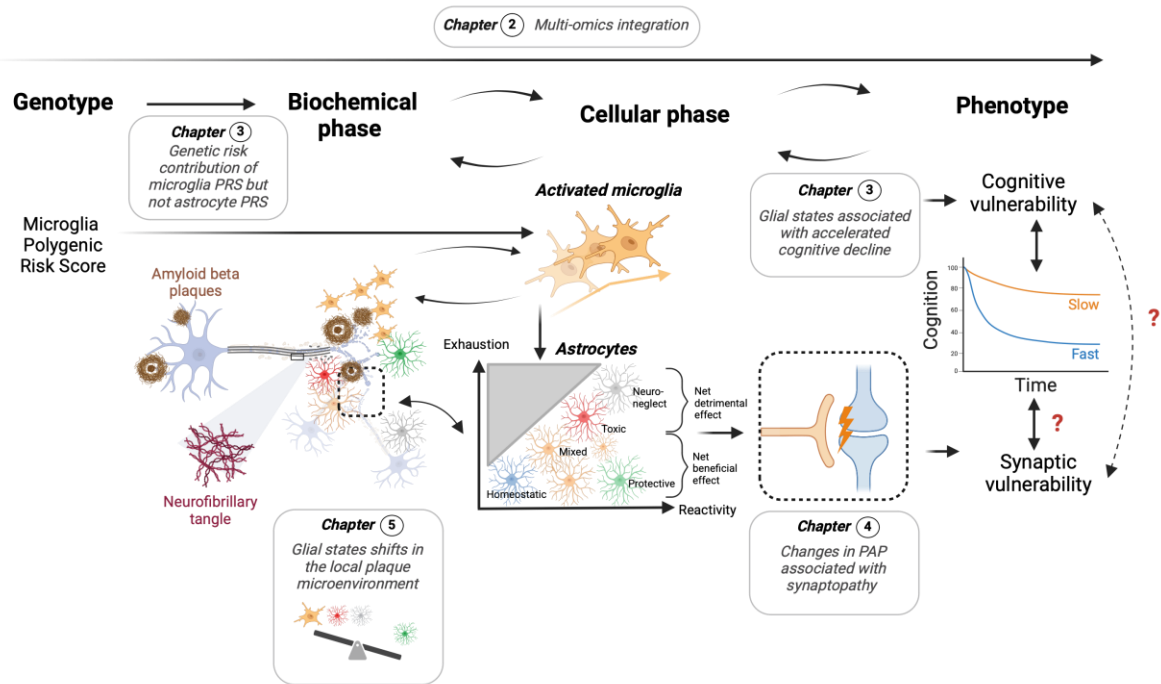


Figure 6-1 Thesis graphical summary

Multi-scale integration of genotype, biochemical, cellular, and clinical phases in AD. The genotype phase highlights the role of a microglial PRS in modulating microglial activation and its downstream effects. The biochemical phase includes the formation of amyloid beta plaques and neurofibrillary tangles, initiating pathological cascades. The cellular phase features activated microglia and astrocytes, with distinct subpopulations. The phenotypic phase outlines synaptic vulnerability and cognitive decline, emphasising the differential trajectories of slow and fast cognitive decline. Each chapter focuses on specific aspects: Chapter 3 explores genetic risk factors, glial activation and link to cognitive decline; Chapter 4 examines reactive astrocytes and their relationship to synaptopathy; Chapter 5 investigates spatial glial responses in amyloid plaque niches. Bidirectional arrows depict feedback loops between phases, and the cognition vs. time graph illustrates the impact of these processes on disease progression. Figure generated via [Biorender](#).

It is now widely accepted that AD is not a linear process driven solely by A β deposition, as initially proposed in the controversial amyloid cascade hypothesis⁴. Instead, it is a dynamic, multifaceted disease involving interconnected biochemical, cellular, and clinical stages that evolve over many years and is preceded by a long prodromal phase¹⁸. Still, the biochemical phase of AD is marked by the accumulation of A β into plaques and neurofibrillary tangles, initiating a cascade of molecular changes leading to neurodegeneration^{4,24}. However, recent attention has shifted towards the cellular phase, particularly the activation of glial cells such as astrocytes and microglia. These cells are key players in the brain's response to pathological stress and mediate both protective and detrimental effects. Advanced molecular profiling techniques, such as single-cell RNA sequencing (scRNA-seq), have revealed that these cellular responses are “multipronged” and exhibit “complex dynamics”^{29,40}. Multi-omics and integrative analyses now offer new opportunities for mapping a more comprehensive molecular profile of the cellular changes that occur in AD.

This PhD research followed a bottom-up methodology, bridging the genotype to the phenotype. It mainly focused on linking these cellular dynamics to molecular and clinical phenotypes by integrating multi-omics data to investigate how microglial and astrocyte activation impact synaptopathy and differential cognitive decline in AD (**Figure 6-1**). In this concluding chapter, I summarise the key findings, implications, and contributions to the field of AD research. I also highlight some limitations of my work, outstanding questions, translational applications and future research perspectives.

6.1 Multi-omics integration in AD: key research outcomes

6.1.1 Integration scenarios encountered in this thesis

The Latin root of “integrate” is *integrare*, which means “to make whole”. In multi-omics research, various omics layers — genomics, transcriptomics, proteomics, etc. can be leveraged to create a more complete picture of biological systems. When analysed jointly, they bring both unique and complementary insights, by unravelling interactions, regulatory mechanisms, and biological pathways that aren't necessarily apparent in isolation. With an integrated molecular footprint, researchers can better understand disease phenotypes and sources of disease heterogeneity. In **Chapter 1**, I introduced vertical integration as an integration strategy for the

joint analysis of multiple omic layers derived from the same donor. This differs from horizontal integration, mainly used to integrate several cohorts into one meta-cohort. In this thesis, I did not horizontally integrate cohorts and merely used external cohorts to validate results derived from the Matthews group's internal cohort. As summarised below, various vertical integration methods were used throughout this thesis to unravel the complexities of AD.

In **Chapter 2**, I developed *Omix*, an R software package designed to integrate bulk transcriptomic and proteomic data¹²⁰. It streamlines the mapping of gene expression and protein activity networks to reveal key regulatory nodes and biological pathways. This is particularly significant because, to date, no specialised integration workflow for bulk transcriptomics and proteomics, the most frequently recorded omics combination in non-cancer-focused studies⁸⁵, has been documented. *Omix* is a wrapper package that offers a comprehensive range of processing, visualisation functions, and state-of-the-art integrative models. Altogether, this is to ensure speed, scalability, and reproducibility in joint transcriptomics-proteomics analyses. The pipeline was validated using real-world AD data, and findings underscore the usefulness of *Omix* in elucidating disease mechanisms and regulatory pathways and identifying both established and novel therapeutic targets. More specifically, disease signatures derived from multi-omics integration were compared against single omic transcriptomics and proteomics signatures and appeared more biologically relevant, predictive of AD global pathology and robust to down-sampling. In a case study, I also demonstrated *Omix*'s capability to generate a pseudotemporal disease profile, providing valuable insights into potential regulatory pathways by identifying transcription factor proteins, mainly STAT1, STAT3 and CEBPB, that regulate a microglia gene network in the disease state. I validated these findings using single nuclei ATAC-seq data to confirm that the identified transcription factor proteins have differentially accessible chromatin regions at their target sites, identified through motif analysis, supporting their active regulatory roles in AD. Overall, *Omix* is an R package that can be applied agnostically to integrate bulk transcriptomics and untargeted proteomics datasets. *Omix* empowers both experts and non-experts to harness the full potential of multi-modal data, in a reproducible fashion, thereby enabling rapid advances in integrative multi-omics research.

In **Chapter 3**, I used the *Omix* pipeline to integrate bulk transcriptomics-proteomics data again and build AD-specific disease modules. I prioritised three modules, F14_UP_ME1,

F14_UP_ME2 and F11_DOWN_ME2, enriched for microglia, astrocyte and neuronal cell types, respectively. These modules served as building blocks and were further integrated downstream in various ways:

- 1. Integration of bulk modules and single-cell substate-specific profiles:** I used snRNA-seq data from matching donors to find subpopulations of microglia, astrocytes and neurons and showed that these transcriptomics-proteomics bulk modules were expressed in specific subpopulations.
- 2. Integration of bulk modules and genetic risk:** I carried out genetic risk enrichment tests based on AD GWAS, quantified astrocyte and microglia-specific PRS, and showed that the latter PRS directly modulates microglial activation, as quantified by the transcriptomic-proteomic module F14_UP_ME1. The astrocyte PRS doesn't directly modulate F14_UP_ME2 astrocyte reactivity, though it was found to be associated with AD neuropathology.
- 3. Integration of bulk modules with CSF or plasma proteomic modules:** I studied the overlap of these modules with CSF and plasma modules from Dammer et al.¹⁸² to help designate candidate biomarkers.
- 4. Unsupervised integration to find molecular subgroups:** I used unsupervised multi-omics clustering to find molecular subgroups in early AD (Braak III-IV), representing a transitional stage in disease progression (i.e. that is therefore more likely to reveal molecular heterogeneity). The two molecular subgroups showed differential profiles of glial activation and synaptic dysfunction.
- 5. Integration of omics with retrospective longitudinal trajectories:** Upon retrospective lifetime assessment of cognitive trajectories, the molecular subgroups showed no difference in A β or Tau, though they corresponded with slow and fast trajectories of cognitive decline. In other words, unsupervised clustering without prior knowledge of clinical groups unbiasedly identified relevant clinical phenotypes based on molecular footprinting.

In **Chapter 4**, I used *Omix* in a different integration use case to integrate synaptic and cytosolic fractions of bulk synapse proteomics measurements. This strategy aimed to disentangle AD mechanisms localised to the synapse and contextualise their interactions within the broader cellular environment. Multiple brain regions representing different stages of AD pathology

were also integrated into the model, ultimately mapping a spatio-pseudotemporal profile of synaptic proteomic changes in AD. This work re-emphasised astrocytic contributions to synaptopathy, as initially proposed in **Chapter 3**, now through the lenses of synaptic proteomics.

In **Chapter 5**, I extended these insights by integrating spatial proteomics derived from IMC with the snRNA-seq glial subpopulations defined in **Chapter 3**. This allowed me to validate the glial states detected in bulk and single-nuclei data and further spatially resolve them according to non-neuritic and neuritic amyloid plaques. Linking back to the astrocytic and synaptic insights from **Chapter 4**, I observed spatial patterns of glial activation in proximity to plaques, with more advanced niches showing a depletion of neuroprotective astrocytes and enrichment of neurotoxic astrocytes.

6.1.2 Molecular characterisation of glial activation in AD

In the landscape of AD research, the cellular phase of the disease has increasingly become the focus, particularly the role of glial cells such as microglia and astrocytes. Most AD risk genes revealed via GWAS studies, have been shown to converge on microglial response pathways, including *TREM2*, *BIN1*, or *PICALM*. *APOE* is the biggest risk factor and is mostly expressed by astrocytes. Microglia and astrocytes are at the forefront of the brain's response to neuropathology, and their molecular profiling through scRNA-seq and snRNA-seq has yielded substantial advancements in our understanding of AD pathogenesis. Pioneering single-cell studies, explored in **Chapters 3** and **4**, previously identified diverse microglial and astrocyte phenotypes that become activated during AD progression. Notably, the cellular phase appears central in determining the onset of clinical symptoms, as the failure of glial cells to maintain homeostasis may drive the transition from cellular dysfunction to cognitive decline^{19,20}. More recent findings comparing AD and asymptomatic AD patients with equivalent neuropathological loads of A β plaques and tangles have highlighted striking differences in glial activation^{19,46–48}, providing evidence that reactive gliosis can no longer be considered a secondary event caused by neurodegeneration, but rather as an active player in AD pathogenesis¹⁰. Differential glial activation may also partly explain some of the heterogeneity in clinical outcomes, including variations in the rate of disease progression, such as fast or slow

progression, as recently suggested⁴⁹. Nevertheless, the role of glial activation and its contribution to synaptic and neuronal damage and cognitive decline remains unclear. The research contributions in this thesis re-emphasise some of the current biological theories on glial activation but also provide a human-focused genotype-to-phenotype framework that integrate multi-omics data to uncover molecular and spatial mechanisms of glial activation in AD. My results support the hypothesis that glial activation contributes to cognitive decline through its effect on synaptopathy, establishing a glia-cognition link that may underlie the clinical heterogeneity in AD.

6.1.2.1 Early microglial activation precedes neurotoxic astrocyte response in AD

It is widely acknowledged that microglia activation can be detrimental, contributing to synaptic loss, altering neuronal network function through changes in TNF- α , C1q and IL-1 α , and causing early neuronal dysfunction³². Still, the human microglial response to A β pathology is complex, as exemplified by the work of Mancuso et al., which describes the multitude of microglial states, ranging from DAM to HLA to IRM. A distinctive pro-inflammatory cytokine and chemokine (CRM) response is mounted against oligomeric A β , known to appear earlier than amyloid plaques. In **Chapter 3**, while I identified subpopulations specific to these states, including homeostatic and CRM clusters (Micro0, Micro4), the Micro2 cluster, which aligned with the prioritised bulk microglia module, exhibited concurrent enrichment of DAM, HLA, and TRM genes. Contrary to findings suggesting a protective role for the microglia HLA state²⁹, this thesis' data indicate otherwise, possibly due to the mixed DAM/HLA composition in the Micro2 subpopulation. As a novel angle, I demonstrated the distinct contributions of transcriptomics and proteomics in mediating AD genetic risk in microglia and astrocytes, with transcriptomics revealing several transcription factors regulating the microglial network (CEBPB, STAT1), while the astrocyte network may be regulated predominantly post-transcriptionally. Key pathways such as complement activation and synaptic pruning emerged as central to the microglial response, as discussed in **Chapter 3** and corroborating previous research³²⁰. In spatial analyses from **Chapter 5**, C1q⁺ microglia in the A β plaque microenvironment predominantly interacted with a neurotoxic mixed reactive/exhausted astrocyte phenotype, re-emphasising the role of activated microglia in inducing this neurotoxic reactive astrocyte phenotype³². Furthermore, examining TREM2 loss of function in **Chapter 5**, in individuals carrying the TREM2^{R47H} or TREM2^{R62H} variants, revealed increased AT8⁺ tau

pathology in plaques, underscoring TREM2's beneficial role in restricting A β plaque growth and limiting plaque-associated neuronal damage^{316,321}.

6.1.2.2 A reactive-exhaustion framework for astrocyte response in AD

Reactive astrocytes, induced by pro-inflammatory signals from activated microglia, may exacerbate oxidative stress and contribute to neuronal death, as Liddelow et al. initially suggested³². More recently, Serrano-Pozo et al. were recently the first group to propose that an additional “exhausted” astrocyte state might deviate from the traditional homeostatic-to-reactive trajectory, reflecting this maladaptive response, which likely results from chronic exposure to neuropathology⁴⁰. At the snRNA-seq level, this thesis identified two functionally distinct reactive astrocyte subpopulations: a classical reactive cluster with a neuroprotective-net effect (Astro2&5) and a mixed reactive/exhausted cluster with a neurotoxic-net effect (Astro3&6). The most reliable markers for astrocyte reactivity, including GFAP, HSBP1, CLU, CD44, C3 or APOE, seem to overlap in multiple literature gene sets, entailing that they probably confer the strongest reactivity signal. On the other hand, key pathways of exhausted astrocytes in this thesis' analyses and the literature consistently showcased lipid dysregulations and metallothionein. However, when it comes to designating groups of markers, analyses fundamentally differ in resolutions, disease stages, and brain regions, making cross-comparisons challenging. Thereby, no single marker reliably distinguishes these subtypes, and this highlights a spectrum of states in the reactivity-exhaustion space rather than fully dichotomised phenotypes. I propose a reactive-exhaustion framework, wherein the balance of co-existing “neurotoxic-like” and “neuroprotective-like” astrocytes shifts local glial-glia and glial-neuronal interactions, ultimately determining the functional outcome (**Figure 6-1**). This framework accommodates more complex characterisation than the binary reactivity profiles typically observed in acute models³⁹. Each analysis most likely characterises a different point in the reactivity-exhaustion space, moving from adaptive, neuroprotective responses to maladaptive, neurotoxic dysfunction, ultimately driving disease progression and shaping the local pathological environment.

6.1.2.3 Glial subpopulations correlate with synaptic dysfunction and accelerated cognitive decline in early AD

This was apparent at the *phenotype level* in **Chapter 3**, where both neuroprotective-like F11_UP_ME1 and neurotoxic-like F14_UP_ME2 astrocyte bulk modules were up-regulated in the fast-progressing cognitive decline subgroup, reflecting the co-existence of states and potential compensatory mechanisms at play. Proteomic markers of astrocytic response appear to be most strongly associated with a molecular subtype of early AD patients who experience accelerated cognitive decline, which is characteristic of a neurotoxic response. This suggests that in the disease sequence, neurotoxic-like astrocyte reactivity may follow microglial activation and is a determinant of clinical symptom progression. This also sheds light on synaptic dysfunction as the potential link between glial activation and cognitive decline. Notably, metallothionein, lipid and inflammatory signalling, and extracellular matrix-related pathways, a hallmark of exhausted astrocytes, were also characteristic of the neurotoxic-like astrocyte signature (MT1F, TGFB1, PLEC).

6.1.2.4 Synapse loss links astrocyte dysfunction to cognitive decline in AD

Synapse loss is the strongest neuropathological correlate of cognitive impairment in AD^{53–56}. Therefore, the synapse is a critical substrate linking cognitive symptoms and proteinopathy in AD, making it a promising target for novel therapies. In parallel, astrocytes are the gatekeepers of synaptic function via the tripartite synapse, playing a role in synapse formation, maintenance, and elimination. Looking at the *synapse level* in **Chapter 4**, I characterised a synaptic proteomic cascade of events leading to synapse loss alongside an associated synaptotoxic astrocyte proteomic signature. Briefly, I characterised regional molecular pathology at the synapse level and revealed differential excitatory synapse subtype vulnerability in AD. This approach uncovered that differential synapse subtype loss in AD is associated with specific astrocytic proteins, particularly those localised to peri-synaptic astrocytic processes. These processes were enriched in complement activation and endocytosis, that are pathways implicated in synaptic pruning and synapse regulation. This corroborates recent research that found collaborative partnerships between microglia and astrocytes in the ingestion of synapses in AD^{44,45} which has also been proposed previously³²⁰. Other pathways unravelled dysregulations in astrocyte-neuron interactions, inflammatory cytokine signalling,

regulation of neuron death, and calcium signalling. Altogether, my findings provide evidence for non-cell autonomous mechanisms of synaptic dysfunction driven by astrocytes, aligning with the notion that these cells may transition into an “exhausted”⁴⁰, “neuro-neglect”¹⁹³ state.

6.1.2.5 Spatial heterogeneity reveals a maladaptive glial response from neurotoxic astrocytes in plaque niches

In **Chapter 5**, I explored the *spatial heterogeneity* of microglial and astrocytic states within AD’s neuropathological microenvironment, specifically around amyloid plaques. Using IMC, I mapped glial activation relative to non-neuritic and neuritic amyloid plaques and Tau⁺ neurons. The analysis confirmed the spatial organisation of neuroprotective and neurotoxic-like astrocytes, showing the enrichment of neurotoxic-like astrocytes and activated microglia within the plaque niche. I also observed a relative shift in the local microenvironment toward astrocytic neurotoxicity, particularly pronounced in the presence of the *TREM2*^{R47H} variant and with increased A β plaque severity, ultimately disrupting beneficial neuroprotective-like astrocyte-to-microglia interactions. Neurotoxic-like astrocytes found in the plaque niche closely mimicked the bulbous morphology characterised previously⁴⁵, which is known to have a decreased astrocyte-synapse coverage, upregulate expression of MFGE8 and promote microglia synapse engulfment. The astrocyte module from synaptic proteomics in **Chapter 4**, was p62/SQSTM1⁺, which also appears to be characteristic of bulbous astrocytes. Further work should validate the presence of bulbous PAPs at the synapse and their effect on synapse loss in human tissue.

Together, my results connect glial activation to synaptopathy and cognitive decline in AD and highlight roles for microglial activation in the genesis of AD and later astrocyte activation as a potential determinant of clinical symptom progression. The provided evidence in this thesis suggests that neurotoxic-like astrocytes, which are predominantly located within A β plaques, could drive synapse loss at these pathological sites, ultimately contributing to accelerated cognitive decline. This opens up new avenues of research such that targeting glia, particularly astrocytes, could mitigate cognitive impairment. A specific area for future investigation would be testing whether targeting peri-synaptic astrocyte processes could decelerate or attenuate synapse dysfunction and synapse loss and slow cognitive decline.

6.1.3 Bridging molecular mechanisms and phenotype through resilience and vulnerability

A core theme in this thesis is the broad concepts of “resilience and vulnerability”, explored at both cognitive and synaptic levels. Resilience can be defined in various ways, but I adopted the framework from Mathys et al.¹⁸³, where resilience and vulnerability are determined by the direction of the difference between observed and expected values of an outcome Y. This Y value can be anything from a cognitive score (*Chapter 3*), a synaptic density score (*Chapter 4*), or any other measurement reflecting disease progression.

In *Chapter 3*, I linked **cognitive vulnerability** to reactive astrocytes and synaptopathy, with reactive microglia showing no such association. I also found in *Chapter 4* that **synaptic vulnerability** was associated with neurotoxic peri-synaptic astrocyte processes. Bringing these findings together raises the possibility that synaptic integrity may at least partly explain cognitive vulnerability. In AD, astrocytic dysfunction could lower synaptic plasticity, making the brain more susceptible to damage and accelerating cognitive decline. This leads to the broader implication that synaptopathy might be a critical link between AD’s cellular and clinical phases. Although I couldn’t directly assess cognitive and synaptic resilience/vulnerability in the same samples in this thesis, future work should further characterise how these resilience mechanisms mechanistically influence one another. Some groups are already assessing the relationship between the rate of synapse loss and cognitive trajectories using SV2A PET imaging to monitor synaptic density in vivo³²², and this could be further complemented by fluid biomarkers to monitor glial activation (see *section 3.3.7*). Ultimately, preserving or enhancing synaptic resilience via astrocytes may offer a novel targeted therapeutic route to slow down cognitive decline in AD. Another open-ended question revolves around selective neuronal vulnerability and whether this is directly linked to the identified vulnerable synapse subtypes.

6.2 Overall challenges

Each chapter discussed study-specific challenges and limitations. Still, here, I want to take a step back, reflect on the overarching challenges and limitations I repeatedly encountered throughout this thesis, and explore potential avenues to overcome these in the future.

6.2.1 Bulk, single nuclei, spatial, temporal: reflecting on multi-level resolutions challenges

One of the primary challenges was integrating data from different levels of resolution, from bulk to single cell, to spatial omics. Bulk omics data, including transcriptomics and proteomics provide an overview but lack cell-type specificity, potentially obscuring signals from less abundant cell types. SnRNA-seq offers cell-level resolution but has inherent limitations, including sparsity and biases compared to whole cell transcriptomics, and a reduced representation of glial inflammatory response^{323,324}. Moreover, even at the single-cell resolution, findings from different cohorts can be challenging to compare without re-integrating and re-analysing the data, specifically when dealing with cellular subpopulations. In this thesis, I repeatedly compared gene set signatures from the literature and often found incomplete overlap. This inconsistency arose from differences in cluster resolution, reference comparisons (e.g., cluster vs. all or differentially expressed genes per cluster), and variations in brain regions and cohort compositions, which are challenging to fully account for. The spatial resolution explored via IMC in **Chapter 5** added another layer of complexity. This technology can combine a relatively high number of markers, but these had to be selected *a priori* (and before I began this thesis, which could have influenced the choice of markers differently), which limited the scope of validation given the findings from untargeted omics techniques in the preceding chapters. The selected markers were based on well-validated literature, including established cell type and pathology markers; however, this approach limited the ability to explore more novel, state-associated markers identified later in the project. The technology was in its infancy at the start of this thesis, and nowadays, more powerful spatial omics techniques have been more readily tested on human brain samples²⁸⁷. Going forward, untargeted single-cell spatial multi-omics has the potential to combine all resolutions in a single experimentation, thereby reducing noise while capturing cellular heterogeneity and interactions³²⁵. To comment on the last level of resolution explored in this thesis: time. I approached this from a pseudotemporal angle, as *post mortem* samples can only inform on a single time point. Cross-sectional analyses suffered from the lack of mid-pathology stage tissues in the internal cohort, which may have impacted the generalisability of findings. I tried to overcome this limitation by retrospectively linking *post mortem* omic measurements to lifetime longitudinal cognitive profiles and also leveraging the stereotypical spatiotemporal progression of AD neuropathology by using several brain regions. Nevertheless, collecting

prospective biofluids from living patients alongside cognitive measurements would be the most appropriate way forward.

6.2.2 Chain of inferences and causality

Human *post mortem* brain samples are invaluable in providing multi-omics and neuropathological information on health and disease. However, these often lack lifetime phenotype data, limiting their usability for translational therapeutic discovery. To bypass this problem, this thesis used an internal cohort from the UK DRI Multi-Omics Atlas project (map-ad.org) as the primary discovery engine, but projected results for replication in external AD cohorts that are better phenotypically characterised, which is a common approach for linking cohorts^{150,326}. Precisely, the main AD-related modules and cellular subpopulations were characterised in the internal cohort and were then projected and replicated in the MSBB and ROSMAP cohorts. The hdWGCNA, a popular tool for network analysis and used in **Chapter 3**, was conceived for this specific intent, namely “rather than building a new co-expression network from scratch in a new dataset, one can take the modules from one dataset and project them into the new dataset”¹⁷². This implied recomputing eigenvalues from the same features present in **Chapter 3** discovery modules. Whilst this may be seen as a “chain of inferences”, where downstream analyses following integration involved sequential steps, such as identifying AD-related subtypes, gene modules, and risk genes, with each step building on the outputs of the previous one, this layered approach is a core strength of this thesis. It enabled the integration of multiple AD cohorts and diverse omics and cognitive measurements, making the impact of the integrated datasets greater than the sum of their individual contributions. Moreover, by applying the machine learning model trained on AD disease modules in the internal cohort to predict pseudotime in the ROSMAP cohort, I found that fast-progressing patients in ROSMAP had higher pseudotime, providing evidence that the disease modules identified in the internal cohort are not only transferable but also robustly predictive of disease progression in an independent cohort. The trained model has been made available to the research community to enhance the reproducibility of AD research. In the future, this model may be used to harmonise data from human *post mortem* omic measurements to enable cross-cohort comparisons and facilitate the identification of shared molecular mechanisms underlying disease progression.

While the sequential integration strategy used here is a powerful framework for hypothesis generation and cross-cohort validation, it cannot fully address questions of causality. A major limitation is the possibility of reverse causation, where observed AD-related mechanisms may be consequences rather than drivers of disease. Although this thesis presents converging evidence implicating astrocytes in synaptopathy and cognitive decline, future studies should incorporate causal inference methods to robustly validate these findings. For instance, Mendelian randomisation (MR) methods, using genetic variants as instrumental variables, have recently been applied in single-cell frameworks, further expanding our knowledge of cell-type-specific causal genes in AD³²⁷. Extending this causal framework to focus on AD-related biological mechanisms, particularly in the form of multi-omics modules like in **Chapter 3**, rather than single genes, could provide a more comprehensive view of disease causality. In future work, I will implement a novel cell subtype-specific “EigenQTL” mapping approach, measuring expression quantitative trait loci (eQTL) on candidate disease modules’ eigenvalues (gene and protein networks) rather than individual genes, enabling investigation of causal mechanisms at the pathway level in AD. This will be achieved by computing module eigenvalues in astrocyte and microglia subtypes of interest through pseudobulking. A key consideration will be the potential for horizontal pleiotropy as SNPs may affect the outcome through multiple pathways rather than exclusively through the target module, which may in turn violate the exclusion restriction assumption. To mitigate this, I will select SNPs in the cis window ($\pm 1\text{Mb}$) of genes that not only drive the module’s genetic risk, as identified in previous GWAS studies but also serve as key network nodes, such as those highlighted in **Chapter 3** (section 3.3.5). Follow-up colocalisation analyses (SNP-module for the trait and SNP-AD status as the outcome) will help validate instrumental variables (IV) that influence coordinated networks based on colocalisation posterior probability (PP.H4 or the posterior probability that AD and the module share a single causal variant). IV-module-AD relationships could then be further tested using robust MR methods and thereby robustly identify causal modules.

6.3 Translational applications

6.3.1 Patient stratification and monitoring

Novel clinical decision tools hold the promise of uncovering patient subgroups and ultimately improve the design of clinical trials — essential for new treatments to see light. Fluid

biomarkers, including plasma and CSF, hold promise for predicting disease progression, stratifying patients and monitoring⁶⁴. In **Chapter 3**, by overlapping CSF and plasma proteomics modules from Dammer *et al.*¹⁸² with this thesis' modules, I proposed candidate biomarkers for identifying patients at higher risk for rapid cognitive decline. The neuro-glia score, combining GAP43, NPY, C3, CD14, and AZGP1, demonstrated significant improvements in distinguishing vulnerable and resilient AD cases compared to predictions using A β 42, tTau, pTau, age and sex alone (AUC=0.88 vs. 0.84 and 0.83 vs. 0.76). In summary, I empirically showed that the rate of cognitive decline relates to the level of cognitive resilience or vulnerability, which the proposed markers can predict in addition to A β 42, tTau, pTau, age and sex. For asymptomatic AD individuals experiencing an absence of cognitive decline despite the burden of neuropathology, this predictive model could potentially be a powerful tool to predict future progression. I will analyse individuals' cognitive trajectories as the data becomes available to assess whether individuals predicted to be resilient are more likely to remain stable than progress into more severe forms of AD. If so, this has the potential to compete with existing predictive prognostic models, such as one trained on MRI grey matter atrophy, demographics and cognitive tests, that can robustly predict whether patients at early disease stages (MCI) will remain stable or progress to AD (AUC=0.84)³²⁸. Ultimately, this could enable early identification of high-risk individuals, facilitating improved monitoring, timely interventions, and targeted lifestyle modifications. Large biobanks could also provide valuable opportunities for external validation of the proposed biomarkers and their predictive utility across more diverse populations.

Still, lumbar puncture required for CSF sampling is highly invasive, which limits its widespread clinical adoption. Emerging MRI-based glial markers, performed *ex vivo* on tissue, have recently successfully differentiated between symptomatic and asymptomatic AD⁴⁷. The next challenge will be to develop *in vivo* markers to detect early changes in glial phenotypes and early neuronal and glial vulnerability¹⁹. Plasma-based biomarkers present another avenue with broader clinical adoption potential, but so far, they have primarily focused on detecting A β (A β 42/A β 40) and tau (pTau217, pTau23, tTau). Combining these with glial and synaptic dysfunction markers holds promise. Complement proteins, highlighted in **Chapter 3**, are suitable candidates. Finally, cell-free DNA (cfDNA)-based liquid biopsies are now actively

being explored as an alternative/complement to plasma proteomics, mainly for their potential to inform on brain methylation patterns from the blood and in a cell-type specific manner³²⁹.

6.3.2 Novel therapies

Emerging disease-modifying therapies, such as Lecanemab and Donanemab, represent critical advancements in AD treatment. These amyloid-targeting antibodies aim to prevent aggregation and enhance the clearance of amyloid beta. They are effective in the biochemical phase of the disease and show promise for primary and secondary prevention. However, these therapies largely overlook glial activation's role in the cellular phase, representing a novel intervention avenue.

Microglia exhibit stage-specific roles in AD. Under normal conditions, they contribute to amyloid clearance through phagocytosis. However, their role in early AD remains debated: while traditionally viewed as protective, recent studies have suggested that microglia may contribute to plaque initiation and seeding in early stages of pathology³³⁰. As the disease progresses, they transition to a protective role by compacting amyloid plaques, but in later stages, their overactivation may drive harmful neuroinflammation. TREM2 agonists, such as AL002, until recently in phase 2 clinical trial (NCT04592874)³³¹, aimed to leverage this protective role of plaque compaction in early AD, by targeting individuals in early AD with mild cognitive impairment (MMSE ≥ 20). However, in November 2024, the trial returned unconvincing results. Our findings provide a potential molecular explanation for this failure. While microglial responses may be protective in early pathology, this thesis' data indicated that strategies to downregulate microglial responses may prevent further damage in mid-pathology stages (Braak III-IV), at a stage where microglial overactivation most likely exacerbates neuroinflammation. Specifically, I observed that resilient individuals with MMSE ≥ 20 exhibited a significantly decreased microglial response compared to Braak 0-II, whereas vulnerable individuals maintained an elevated response. This implies that inhibiting the microglial module at mid-pathology stages might be a more effective therapeutic approach (*Supplementary figure 6-1*). It is therefore possible that the trial enrolled individuals with MMSE ≥ 20 who had already accumulated significant neuropathology, as no enrolment criteria directly assessed the extent of underlying pathology³³¹.

I identified key transcription factors, CEBPB³³² and STAT3³³³, as promising molecular targets to regulate this module. Targeting STAT3 through small molecules offers a potential avenue to reduce neuroinflammation mediated by this pathway. However, selective targeting to avoid off-target effects in non-microglial cells or unaffected brain regions will be crucial for clinical translation.

Another therapeutic route to modulate the cellular phase could be to selectively target astrocytes to restore their neuroprotective functions and mitigate synaptotoxic effects. Such approaches could enhance synaptic resilience, improve cognitive outcomes, and delay disease progression. Emerging biotechnology companies like [AstronauTx](#) are currently working on such approaches. Leveraging a cellular phenotypic screening platform, Clayton et al. showed that chemical inhibitors of HDAC3 blocked pathological astrocyte reactivity²²⁴, opening avenues for additional pharmacological interventions to restore astrocytic homeostasis. Drug repurposing is also an active area of research, whereby existing cancer drugs, such as those targeting Ezrin phosphorylation, a target identified in section **Chapter 4** section 4.3.3, are being evaluated for their potential to mitigate astrocytic dysfunction. Indeed, targeting ezrin-mediated PAP plasticity might help regulate astrocytic responses to glutamate and provide neuroprotection²⁶³. Finally, gene-silencing and activation approaches, combined with pharmacological modulators, offer a promising avenue to fine-tune astrocytic responses and have so far only been explored in animal models:

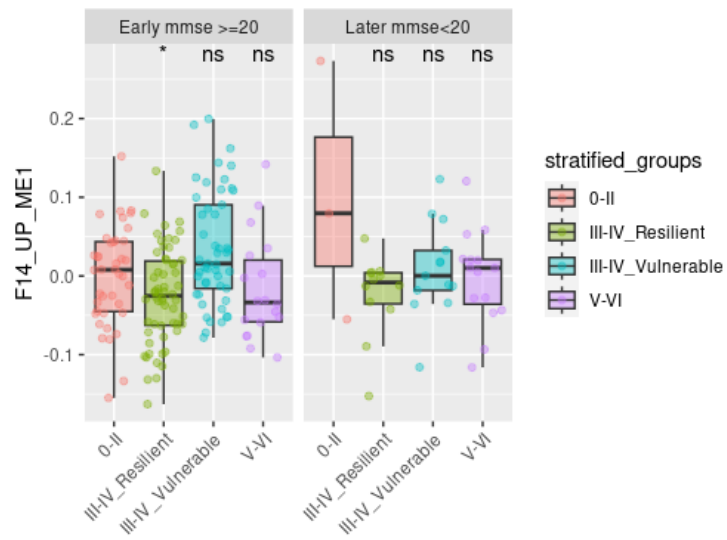
- **Activation of neuroprotective genes:** Enhancing the expression of genes that promote astrocyte-mediated neuroprotection offers a compelling approach to mitigating AD pathology. For example, astrocyte-secreted CLU has been shown to improve excitatory synaptic transmission and reduced A β accumulation in mouse models¹⁹⁴. AAV-mediated astroglial overexpression of *CLU* using astrocyte-specific promoters could serve as a therapeutic strategy to preserve synaptic resilience. This approach could be further explored with small activating RNAs (saRNAs), such as those developed by [MinaTX](#), which selectively boost target gene expression. In the future, saRNAs could be designed to upregulate any astrocytic neuroprotective genes to enhance astrocyte-mediated synaptic support.

- **Silencing genes in maladaptive pathways:** Suppressing pathways that drive astrocytic dysfunction or exacerbate neuroinflammation is another possible route. Targeting the calcineurin/NFAT signalling pathway in astrocytes has shown promise in improving cognitive and synaptic function in AD mouse models³³⁴. AAV vectors carrying the peptide VIVIT, which interferes with calcineurin/NFAT signalling, were delivered under the control of the astrocyte-specific Gfa2 promoter, leading to reduced astrocyte reactivity, improved synaptic function, and lower amyloid levels in APP/PS1 mice. Similarly, researchers have used astrocyte-specific RNA interference (RNAi) systems with shRNAs to silence *LC3B* and *SQSTM1*³³⁵, implicated in autophagy, and the latter found in synaptotoxic astrocytes from **Chapter 4**. By incorporating astrocyte-specific promoters like gfaABC1D into the AAV vectors, this gene-silencing approach successfully ensured a precise targeting of astrocytes. This gene-silencing approach offers the potential for astrocyte modulation as shRNA sequences can theoretically be tailored to target other genes implicated in astrocyte dysfunction.

6.4 Closing paragraph

This work highlights how integrative multi-omics can improve our understanding of complex neurodegenerative disorders by revealing key mechanisms, identifying new therapeutic targets, and uncovering biomarkers for patient stratification. Precisely, My PhD research has advanced the understanding of glial activation in AD. Findings suggest that glial activation is cell state-specific and enriched in AD risk genes, playing a determinative role in disease progression: early microglial activation precedes severe neuropathology, while neurotoxic astrocytic responses are associated with an accelerated rate of cognitive decline. Altogether, my results support the hypothesis that glial activation, particularly that of neurotoxic astrocytes, contributes to cognitive decline through its effects on synaptopathy, establishing a glia-cognition link that may underlie the observed clinical heterogeneity in AD (**Figure 6-1**). Future validation of proposed candidate biomarkers may help predict clinical progression trajectories, thereby informing clinical trial recruitment and providing new therapeutic routes targeting dysfunctional glial cells near the synapse in AD.

6.5 Supplementary figures



Supplementary figure 6-1 Microglia module eigenvalue stratified by MMSE and Braak groups

Bibliography

1. Jack, C. R. *et al.* Prevalence of Biologically vs Clinically Defined Alzheimer Spectrum Entities Using the National Institute on Aging–Alzheimer’s Association Research Framework. *JAMA Neurol* **76**, 1174 (2019).
2. Glenner, G. G. & Wong, C. W. Alzheimer’s disease: Initial report of the purification and characterization of a novel cerebrovascular amyloid protein. *Biochem Biophys Res Commun* **120**, 885–890 (1984).
3. Brion, J.-P., Passareiro, H., Nunez, J. & Flament-Durand, J. IMMUNOLOGICAL DEMONSTRATION OF TAU PROTEIN INTO NEUROFIBRILLARY TANGLES OF ALZHEIMER’S DISEASE. *Arch. Biol. (Bruxelles)* vol. 95 (1985).
4. Hardy, J. A. & Higgins, G. A. Alzheimer’s disease: The amyloid cascade hypothesis. *Science* vol. 256 184–185 Preprint at <https://doi.org/10.1126/science.1566067> (1992).
5. Kang, J. *et al.* The precursor of Alzheimer’s disease amyloid A4 protein resembles a cell-surface receptor. *Nature* **325**, 733–736 (1987).
6. Levy-Lahad, E. *et al.* A Familial Alzheimer’s Disease Locus on Chromosome 1. *Science* (1979) **269**, 970–973 (1995).
7. De Strooper, B. *et al.* Deficiency of presenilin-1 inhibits the normal cleavage of amyloid precursor protein. *Nature* **391**, 387–390 (1998).
8. Strittmatter, W. J. *et al.* Apolipoprotein E: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. *Proceedings of the National Academy of Sciences* **90**, 1977–1981 (1993).
9. Sierksma, A., Escott-Price, V. & De Strooper, B. Translating genetic risk of Alzheimer’s disease into mechanistic insight and drug targets. *Science* (1979) **370**, 61–66 (2020).
10. Chung, W.-S., Welsh, C. A., Barres, B. A. & Stevens, B. Do glia drive synaptic and cognitive impairment in disease? *Nat Neurosci* **18**, 1539–1545 (2015).
11. Hebert, L. E., Weuve, J., Scherr, P. A. & Evans, D. A. Alzheimer disease in the United States (2010–2050) estimated using the 2010 census. *Neurology* **80**, 1778–1783 (2013).
12. Livingston, G. *et al.* Dementia prevention, intervention, and care: 2024 report of the Lancet standing Commission. *The Lancet* **404**, 572–628 (2024).
13. Corder, E. H. *et al.* Gene Dose of Apolipoprotein E Type 4 Allele and the Risk of Alzheimer’s Disease in Late Onset Families. *Science* (1979) **261**, 921–923 (1993).
14. Jin, S. C. *et al.* Coding variants in TREM2 increase risk for Alzheimer’s disease. *Hum Mol Genet* **23**, 5838–5846 (2014).
15. Escott-Price, V. *et al.* Common polygenic variation enhances risk prediction for Alzheimer’s disease. *Brain* **138**, 3673–3684 (2015).
16. Stevenson-Hoare, J. *et al.* Plasma biomarkers and genetics in the diagnosis and prediction of Alzheimer’s disease. *Brain* **146**, 690–699 (2023).
17. Scheltens, P. *et al.* Alzheimer’s disease. *The Lancet* **397**, 1577–1590 (2021).
18. De Strooper, B. & Karran, E. The Cellular Phase of Alzheimer’s Disease. *Cell* **164**, 603–615 (2016).
19. Taddei, R. N. *et al.* Changes in glial cell phenotypes precede overt neurofibrillary tangle formation, correlate with markers of cortical cell damage, and predict cognitive status of individuals at Braak III-IV stages. *Acta Neuropathol Commun* **10**, 72 (2022).

20. De Sousa, R. A. L. Reactive gliosis in Alzheimer's disease: a crucial role for cognitive impairment and memory loss. *Metab Brain Dis* **37**, 851–857 (2022).
21. Long, J. M. & Holtzman, D. M. Alzheimer Disease: An Update on Pathobiology and Treatment Strategies. *Cell* **179**, 312–339 (2019).
22. Hatami, A., Monjaze, S. & Glabe, C. The Anti-Amyloid- β Monoclonal Antibody 4G8 Recognizes a Generic Sequence-Independent Epitope Associated with α -Synuclein and Islet Amyloid Polypeptide Amyloid Fibrils. *Journal of Alzheimer's Disease* **50**, 517–525 (2016).
23. Kepp, K. P., Robakis, N. K., Høilund-Carlsen, P. F., Sensi, S. L. & Vissel, B. The amyloid cascade hypothesis: an updated critical review. *Brain* **146**, 3969–3990 (2023).
24. Braak, H. & Braak, E. Neuropathological staging of Alzheimer-related changes. *Acta Neuropathol* **82**, 239–259 (1991).
25. Malia, T. J. *et al.* Epitope mapping and structural basis for the recognition of phosphorylated tau by the anti-tau antibody AT8. *Proteins: Structure, Function, and Bioinformatics* **84**, 427–434 (2016).
26. Otvos, L. *et al.* Monoclonal antibody PHF-1 recognizes tau protein phosphorylated at serine residues 396 and 404. *J Neurosci Res* **39**, 669–673 (1994).
27. Hulshof, L. A., van Nuijs, D., Hol, E. M. & Middeldorp, J. The Role of Astrocytes in Synapse Loss in Alzheimer's Disease: A Systematic Review. *Front Cell Neurosci* **16**, (2022).
28. Zlokovic, B. V. Neurovascular pathways to neurodegeneration in Alzheimer's disease and other disorders. *Nat Rev Neurosci* **12**, 723–738 (2011).
29. Mancuso, R. *et al.* Xenografted human microglia display diverse transcriptomic states in response to Alzheimer's disease-related amyloid- β pathology. *Nat Neurosci* (2024) doi:10.1038/s41593-024-01600-y.
30. Penney, J. *et al.* iPSC-derived microglia carrying the TREM2 R47H⁺ mutation are proinflammatory and promote synapse loss. *Glia* **72**, 452–469 (2024).
31. Fancy, N. *et al.* Mechanisms contributing to differential genetic risks for TREM2 R47H and R62H variants in Alzheimer's Disease. doi:10.1101/2022.07.12.22277509.
32. Liddel, S. A. *et al.* Neurotoxic reactive astrocytes are induced by activated microglia. *Nature* **541**, 481–487 (2017).
33. Paolicelli, R. C. *et al.* Defining Microglial States and Nomenclature: A Roadmap to 2030. *Diego Gomez-Nicola* vol. 43 <https://ssrn.com/abstract=4065080> (2022).
34. Liddel, S. A. & Barres, B. A. Reactive Astrocytes: Production, Function, and Therapeutic Potential. *Immunity* **46**, 957–967 (2017).
35. Matusova, Z., Hol, E. M., Pekny, M., Kubista, M. & Valihrach, L. Reactive astrogliosis in the era of single-cell transcriptomics. *Front Cell Neurosci* **17**, (2023).
36. Arranz, A. M. & De Strooper, B. The role of astroglia in Alzheimer's disease: pathophysiology and clinical implications. *Lancet Neurol* **18**, 406–414 (2019).
37. Habib, N. *et al.* Disease-associated astrocytes in Alzheimer's disease and aging. *Nat Neurosci* **23**, 701–706 (2020).
38. Livingston, N. R. *et al.* Relationship between astrocyte reactivity, using novel 11C-BU99008 PET, and glucose metabolism, grey matter volume and amyloid load in cognitively impaired individuals. *Mol Psychiatry* **27**, 2019–2029 (2022).
39. Jiawaji, Z. *et al.* Reactive astrocytes acquire neuroprotective as well as deleterious signatures in response to Tau and A β pathology. *Nat Commun* **13**, 135 (2022).

40. Serrano-Pozo, A. *et al.* Astrocyte transcriptomic changes along the spatiotemporal progression of Alzheimer's disease. *Nat Neurosci* (2024) doi:10.1038/s41593-024-01791-4.
41. Liddel, S. A., Marsh, S. E. & Stevens, B. Microglia and Astrocytes in Disease: Dynamic Duo or Partners in Crime? *Trends Immunol* **41**, 820–835 (2020).
42. Guttikonda, S. R. *et al.* Fully defined human pluripotent stem cell-derived microglia and tri-culture system model C3 production in Alzheimer's disease. *Nat Neurosci* **24**, 343–354 (2021).
43. Mallach, A. *et al.* Microglia-astrocyte crosstalk in the amyloid plaque niche of an Alzheimer's disease mouse model, as revealed by spatial transcriptomics. *Cell Rep* **43**, 114216 (2024).
44. Tzioras, M. *et al.* Human astrocytes and microglia show augmented ingestion of synapses in Alzheimer's disease via MFG-E8. *Cell Rep Med* **4**, 101175 (2023).
45. Sokolova, D. *et al.* Astrocyte-derived MFG-E8 facilitates microglial synapse elimination in Alzheimers disease mouse models. *bioRxiv* (2024) doi:10.1101/2024.08.31.606944.
46. Barroeta-Espar, I. *et al.* Distinct cytokine profiles in human brains resilient to Alzheimer's pathology. *Neurobiol Dis* **121**, 327–337 (2019).
47. Barsoum, S. *et al.* Resiliency to Alzheimer's disease neuropathology can be distinguished from dementia using cortical astrogliosis imaging. *bioRxiv* (2024) doi:10.1101/2024.05.06.592719.
48. Perez-Nievas, B. G. *et al.* Dissecting phenotypic traits linked to human resilience to Alzheimer's pathology. *Brain* **136**, 2510–2526 (2013).
49. Kouri, N. *et al.* Clinicopathologic Heterogeneity and Glial Activation Patterns in Alzheimer Disease. *JAMA Neurol* **81**, 619 (2024).
50. Hesse, R. *et al.* Comparative profiling of the synaptic proteome from Alzheimer's disease patients with focus on the APOE genotype. *Acta Neuropathol Commun* **7**, 214 (2019).
51. Colom-Cadena, M. *et al.* The clinical promise of biomarkers of synapse damage or loss in Alzheimer's disease. *Alzheimers Res Ther* **12**, 21 (2020).
52. Scheff, S. W. & Price, D. A. Synaptic pathology in Alzheimer's disease: a review of ultrastructural studies. *Neurobiol Aging* **24**, 1029–1046 (2003).
53. Bereczki, E. *et al.* Synaptic markers of cognitive decline in neurodegenerative diseases: a proteomic approach. *Brain* **141**, 582–595 (2018).
54. Oh, H. S.-H. *et al.* Synapse protein signatures in cerebrospinal fluid and plasma predict cognitive maintenance versus decline in Alzheimer's disease. *bioRxiv* 2024.07.22.604680 (2024) doi:10.1101/2024.07.22.604680.
55. Terry, R. D. *et al.* Physical basis of cognitive alterations in alzheimer's disease: Synapse loss is the major correlate of cognitive impairment. *Ann Neurol* **30**, 572–580 (1991).
56. DeKosky, S. T. & Scheff, S. W. Synapse loss in frontal cortex biopsies in Alzheimer's disease: Correlation with cognitive severity. *Ann Neurol* **27**, 457–464 (1990).
57. Ray, A., Loghinov, I., Ravindranath, V. & Barth, A. L. Early hippocampal hyperexcitability and synaptic reorganization in mouse models of amyloidosis. *iScience* **27**, 110629 (2024).
58. Targa Dias Anastacio, H., Matosin, N. & Ooi, L. Neuronal hyperexcitability in Alzheimer's disease: what are the drivers behind this aberrant phenotype? *Transl Psychiatry* **12**, 257 (2022).

59. McKhann, G. M. *et al.* The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia* **7**, 263–269 (2011).
60. Pangman, V. C., Sloan, J. & Guse, L. An examination of psychometric properties of the Mini-Mental State Examination and the Standardized Mini-Mental State Examination: Implications for clinical practice. *Applied Nursing Research* **13**, 209–213 (2000).
61. Rockwood, K., Strang, D., MacKnight, C., Downer, R. & Morris, J. C. Interrater Reliability of the Clinical Dementia Rating in a Multicenter Trial. *J Am Geriatr Soc* **48**, 558–559 (2000).
62. Koval, I. *et al.* AD Course Map charts Alzheimer's disease progression. *Sci Rep* **11**, 8020 (2021).
63. Wang, X., Ye, T., Zhou, W. & Zhang, J. Uncovering heterogeneous cognitive trajectories in mild cognitive impairment: a data-driven approach. *Alzheimers Res Ther* **15**, 57 (2023).
64. Tijms, B. M. *et al.* Cerebrospinal fluid proteomics in patients with Alzheimer's disease reveals five molecular subtypes with distinct genetic risk profiles. *Nat Aging* **4**, 33–47 (2024).
65. Eteleeb, A. M. *et al.* Brain high-throughput multi-omics data reveal molecular heterogeneity in Alzheimer's disease. *PLoS Biol* **22**, e3002607 (2024).
66. Ebenau, J. L. *et al.* ATN classification and clinical progression in subjective cognitive decline. *Neurology* **95**, (2020).
67. Subramanian, I., Verma, S., Kumar, S., Jere, A. & Anamika, K. Multi-omics Data Integration, Interpretation, and Its Application. *Bioinformatics and Biology Insights* vol. 14 Preprint at <https://doi.org/10.1177/1177932219899051> (2020).
68. Daunt, P. *et al.* Polygenic Risk Scoring is an Effective Approach to Predict Those Individuals Most Likely to Decline Cognitively Due to Alzheimer's Disease. *J Prev Alzheimers Dis* 1–6 (2020) doi:10.14283/jpad.2020.64.
69. Yang, H.-S. *et al.* Cell-type-specific Alzheimer's disease polygenic risk scores are associated with distinct disease processes in Alzheimer's disease. *Nat Commun* **14**, 7659 (2023).
70. Mathys, H. *et al.* Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332–337 (2019).
71. Muntel, J. *et al.* Surpassing 10 000 identified and quantified proteins in a single run by optimizing current LC-MS instrumentation and data analysis strategy. *Mol Omics* **15**, 348–360 (2019).
72. Wang, Z. *et al.* High-throughput proteomics of nanogram-scale samples with Zeno SWATH MS. *Elife* **11**, (2022).
73. Seyfried, N. T. *et al.* A Multi-network Approach Identifies Protein-Specific Co-expression in Asymptomatic and Symptomatic Alzheimer's Disease. *Cell Syst* **4**, 60–72.e4 (2017).
74. Keenan, A. B. *et al.* ChEA3: transcription factor enrichment analysis by orthogonal omics integration. *Nucleic Acids Res* **47**, W212–W224 (2019).
75. Baysoy, A., Bai, Z., Satija, R. & Fan, R. The technological landscape and applications of single-cell multi-omics. *Nat Rev Mol Cell Biol* **24**, 695–713 (2023).
76. Weinstein, J. N. *et al.* The cancer genome atlas pan-cancer analysis project. *Nat Genet* **45**, 1113–1120 (2013).

77. Zhang, L. *et al.* Deep learning-based multi-omics data integration reveals two prognostic subtypes in high-risk neuroblastoma. *Front Genet* **9**, (2018).
78. Zheng, C. & Xu, R. Molecular subtyping of Alzheimer's disease with consensus non-negative matrix factorization. *PLoS One* **16**, (2021).
79. Duan, R. *et al.* Evaluation and comparison of multi-omics data integration methods for cancer subtyping. *PLoS Comput Biol* **17**, (2021).
80. Schulte-Sasse, R., Budach, S., Hnisz, D. & Marsico, A. Integration of multiomics data with graph convolutional networks to identify new cancer genes and their associated molecular mechanisms. *Nat Mach Intell* **3**, 513–526 (2021).
81. Abbas, Z., Tayara, H. & Chong, K. T. Alzheimer's disease prediction based on continuous feature representation using multi-omics data integration. *Chemometrics and Intelligent Laboratory Systems* **223**, (2022).
82. Allen, M. *et al.* Human whole genome genotype and transcriptome data for Alzheimer's and other neurodegenerative diseases. *Sci Data* **3**, (2016).
83. Wang, M. *et al.* The Mount Sinai cohort of large-scale genomic, transcriptomic and proteomic data in Alzheimer's disease. *Sci Data* **5**, (2018).
84. Jager, P. L. De *et al.* Data descriptor: A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research. *Sci Data* **5**, (2018).
85. Athieniti, E. & Spyrou, G. M. A guide to multi-omics data collection and integration for translational medicine. *Comput Struct Biotechnol J* **21**, 134–149 (2023).
86. Maier, T., Güell, M. & Serrano, L. Correlation of mRNA and protein in complex biological samples. *FEBS Lett* **583**, 3966–3973 (2009).
87. Vogel, C. & Marcotte, E. M. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet* **13**, 227–232 (2012).
88. Xicota, L. *et al.* Multi-omics signature of brain amyloid deposition in asymptomatic individuals at-risk for Alzheimer's disease: The INSIGHT-preAD study. *EBioMedicine* **47**, 518–528 (2019).
89. Xie, L. *et al.* Identification of functionally connected multi-omic biomarkers for Alzheimer's disease using modularity-constrained Lasso. *PLoS One* **15**, (2020).
90. Klein, H. U., Schäfer, M., Bennett, D. A., Schwender, H. & de Jager, P. L. Bayesian integrative analysis of epigenomic and transcriptomic data identifies Alzheimer's disease candidate genes and networks. *PLoS Comput Biol* **16**, (2020).
91. Peña-Bautista, C., Álvarez-Sánchez, L., Cañada-Martínez, A. J., Baquero, M. & Cháfer-Pericás, C. Epigenomics and lipidomics integration in alzheimer disease: Pathways involved in early stages. *Biomedicines* **9**, (2021).
92. Madrid, L. *et al.* Alzheimer's Disease Neuroimaging Initiative (ADNI) *, EADI Consortium, CHARGE Consortium, GERAD Consortium, GR@ACE/DEGESCO Consortium. Carmen Antúnez-Almagro vol. 13 www.aging-us.com (2021).
93. Moon, S. & Lee, H. JDSNMF: Joint deep semi-non-negative matrix factorization for learning integrative representation of molecular signals in Alzheimer's disease. *J Pers Med* **11**, (2021).
94. Clark, C., Dayon, L., Masoodi, M., Bowman, G. L. & Popp, J. An integrative multi-omics approach reveals new central nervous system pathway alterations in Alzheimer's disease. *Alzheimers Res Ther* **13**, (2021).
95. Abdullah, M. N., Wah, Y. B., Majeed, A. B. A., Zakaria, Y. & Shaadan, N. Identification of Blood-Based Multi-Omics Biomarkers for Alzheimer's Disease Using Firth's Logistic Regression. *Pertanika Journal of Social Sciences and Humanities* **30**, 1197–1218 (2022).

96. Park, J. *et al.* Multi-Omics-Based Autophagy-Related Untypical Subtypes in Patients with Cerebral Amyloid Pathology. *Advanced Science* 2201212 (2022) doi:10.1002/advs.202201212.
97. Yang, M. *et al.* Multi-omic integration via similarity network fusion to detect molecular subtypes of ageing. *Brain Commun* **5**, (2023).
98. Kodam, P., Sai Swaroop, R., Pradhan, S. S., Sivaramakrishnan, V. & Vadrevu, R. Integrated multi-omics analysis of Alzheimer's disease shows molecular signatures associated with disease progression and potential therapeutic targets. *Sci Rep* **13**, 3695 (2023).
99. Liu, A., Fernandes, B. S., Citu, C. & Zhao, Z. Unraveling the intercellular communication disruption and key pathways in Alzheimer's disease: an integrative study of single-nucleus transcriptomes and genetic association. *Alzheimers Res Ther* **16**, 3 (2024).
100. Vacher, M., Canovas, R., Laws, S. M. & Doecke, J. D. A comprehensive multi-omics analysis reveals unique signatures to predict Alzheimer's disease. *Frontiers in Bioinformatics* **4**, (2024).
101. Hoadley, K. A. *et al.* Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell* **158**, 929–944 (2014).
102. Wörheide, M. A., Krumsiek, J., Kastenmüller, G. & Arnold, M. Multi-omics integration in biomedical research – A metabolomics-centric review. *Analytica Chimica Acta* vol. 1141 144–162 Preprint at <https://doi.org/10.1016/j.aca.2020.10.038> (2021).
103. Mirza, B. *et al.* Machine learning and integrative analysis of biomedical big data. *Genes* vol. 10 Preprint at <https://doi.org/10.3390/genes10020087> (2019).
104. Wu, C. *et al.* A selective review of multi-level omics data integration using variable selection. *High-Throughput* vol. 8 Preprint at <https://doi.org/10.3390/ht8010004> (2019).
105. Singh, A. *et al.* DIABLO: An integrative approach for identifying key molecular drivers from multi-omics assays. *Bioinformatics* **35**, 3055–3062 (2019).
106. Argelaguet, R. *et al.* MOFA+: A statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biol* **21**, (2020).
107. Zhang. Integrative clustering methods for multi-omics data. *Computational Stats* (2021).
108. Wang, T. *et al.* MOGONET integrates multi-omics data using graph convolutional networks allowing patient classification and biomarker identification. *Nat Commun* **12**, (2021).
109. Bersanelli, M. *et al.* Methods for the integration of multi-omics data: Mathematical aspects. *BMC Bioinformatics* **17**, (2016).
110. Haider, S. & Pal, R. *Send Orders of Reprints at Reprints@benthamscience.Net Integrated Analysis of Transcriptomic and Proteomic Data. Current Genomics* vol. 14 <http://www.ingenuity.com/> (2013).
111. Schwanhäusser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337–342 (2011).
112. Ludvigsen, M. & Honoré, B. Transcriptomics and Proteomics: Integration? in *eLS* 1–7 (Wiley, 2018). doi:10.1002/9780470015902.a0006188.pub2.
113. Tarazona, S., Arzalluz-Luque, A. & Conesa, A. Undisclosed, unmet and neglected challenges in multi-omics studies. *Nat Comput Sci* **1**, 395–402 (2021).

114. Krassowski, M., Das, V., Sahu, S. K. & Misra, B. B. State of the Field in Multi-Omics Research: From Computational Needs to Data Mining and Sharing. *Frontiers in Genetics* vol. 11 Preprint at <https://doi.org/10.3389/fgene.2020.610798> (2020).
115. Ulfenborg, B. Vertical and horizontal integration of multi-omics data with miodin. *BMC Bioinformatics* **20**, 649 (2019).
116. Zoppi, J., Guillaume, J. F., Neunlist, M. & Chaffron, S. MiBiOmics: an interactive web application for multi-omics data exploration and integration. *BMC Bioinformatics* **22**, (2021).
117. Bredikhin, D., Kats, I. & Stegle, O. MUON: multimodal omics analysis framework. *Genome Biol* **23**, 42 (2022).
118. Lu, X. , M. J. , Z. Y. , J. L. , and Y. F. MOVICS: an R package for multi-omics integration and visualization in cancer subtyping. *bioRxiv* (2020).
119. Krassowski, M., Das, V., Sahu, S. K. & Misra, B. B. State of the Field in Multi-Omics Research: From Computational Needs to Data Mining and Sharing. *Frontiers in Genetics* vol. 11 Preprint at <https://doi.org/10.3389/fgene.2020.610798> (2020).
120. Schneegans, E. *et al.* Omix: A Transcriptomics-Proteomics Integration Pipeline. (2023).
121. Ramos, M. *et al.* Software for the Integration of Multiomics Experiments in Bioconductor. *Cancer Res* **77**, e39–e42 (2017).
122. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
123. Čuklina, J. *et al.* Diagnostics and correction of batch effects in large-scale proteomic studies: a tutorial. *Mol Syst Biol* **17**, (2021).
124. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47–e47 (2015).
125. Langfelder, P. & Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, (2008).
126. Lê Cao, K.-A., Boitard, S. & Besse, P. Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. *BMC Bioinformatics* **12**, 253 (2011).
127. Mo, Q. & Shen, R. *IClusterPlus: Integrative Clustering of Multiple Genomic Data Sets*. <http://cran.r-project.org/web/packages/glmnet/index.html> (2018).
128. Akalin, A. *Computational Genomics with R*. (Chapman and Hall/CRC, 2020). doi:10.1201/9780429084317.
129. Mo, Q. *et al.* A fully Bayesian latent variable model for integrative clustering analysis of multi-type omics data. *Biostatistics* **19**, 71–86 (2018).
130. Tenenhaus, A. *et al.* Variable selection for generalized canonical correlation analysis. *Biostatistics* **15**, 569–583 (2014).
131. Lê Cao, K.-A. & Welham, Z. M. *Multivariate Data Integration Using R*. (Chapman and Hall/CRC, Boca Raton, 2021). doi:10.1201/9781003026860.
132. Langfelder, P. & Horvath, S. Eigengene networks for studying the relationships between co-expression modules. *BMC Syst Biol* **1**, 54 (2007).
133. Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013).
134. Skene, N. G. & Grant, S. G. N. Identification of vulnerable cell types in major brain disorders using single cell transcriptomes and expression weighted cell type enrichment. *Front Neurosci* **10**, (2016).

135. Street, K. *et al.* Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477 (2018).
136. Ochoa, D. *et al.* Open Targets Platform: supporting systematic drug–target identification and prioritisation. *Nucleic Acids Res* **49**, D1302–D1310 (2021).
137. Csárdi, G. & Nepusz, T. The igraph software package for complex network research. in (2006).
138. Hodge, R. D. *et al.* Conserved cell types with divergent features in human versus mouse cortex. *Nature* **573**, 61–68 (2019).
139. de Jager, P. L. *et al.* Data descriptor: A multi-omic atlas of the human frontal cortex for aging and Alzheimer’s disease research. *Sci Data* **5**, (2018).
140. Yu, L. *et al.* Targeted brain proteomics uncover multiple pathways to Alzheimer’s dementia. *Ann Neurol* **84**, 78–88 (2018).
141. Wang, X. *et al.* Deciphering cellular transcriptional alterations in Alzheimer’s disease brains. *Mol Neurodegener* **15**, (2020).
142. Stuart, T., Srivastava, A., Madad, S., Lareau, C. A. & Satija, R. Single-cell chromatin state analysis with Signac. *Nat Methods* **18**, 1333–1341 (2021).
143. Bellenguez, C. *et al.* New insights into the genetic etiology of Alzheimer’s disease and related dementias. *Nat Genet* **54**, 412–436 (2022).
144. Johnson, E. C. B. *et al.* Large-scale deep multi-layer analysis of Alzheimer’s disease brain reveals strong proteomic disease-related changes not observed at the RNA level. *Nat Neurosci* **25**, 213–225 (2022).
145. Bai, B. *et al.* Deep Multilayer Brain Proteomics Identifies Molecular Networks in Alzheimer’s Disease Progression. *Neuron* **105**, 975-991.e7 (2020).
146. Chen, L.-L., Fan, Y.-G., Zhao, L.-X., Zhang, Q. & Wang, Z.-Y. The metal ion hypothesis of Alzheimer’s disease and the anti-neuroinflammatory effect of metal chelators. *Bioorg Chem* **131**, 106301 (2023).
147. Juárez-Rebollar, D., Rios, C., Nava-Ruiz, C. & Méndez-Armenta, M. Metallothionein in Brain Disorders. *Oxid Med Cell Longev* **2017**, (2017).
148. Smith, A. M. *et al.* Diverse human astrocyte and microglial transcriptional responses to Alzheimer’s pathology. *Acta Neuropathol* **143**, 75–91 (2022).
149. Quinn, J. P., Kandigian, S. E., Trombetta, B. A., Arnold, S. E. & Carlyle, B. C. VGF as a biomarker and therapeutic target in neurodegenerative and psychiatric diseases. *Brain Commun* **3**, fcab261 (2021).
150. Beckmann, N. D. *et al.* Multiscale causal networks identify VGF as a key regulator of Alzheimer’s disease. *Nat Commun* **11**, (2020).
151. Leng, F. & Edison, P. Neuroinflammation and microglial activation in Alzheimer disease: where do we go from here? *Nature Reviews Neurology* vol. 17 157–172 Preprint at <https://doi.org/10.1038/s41582-020-00435-y> (2021).
152. Millot, P. *et al.* STAT3 inhibition reverses neuroinflammation and A β metabolism induced by systemic inflammation. *Alzheimer’s & Dementia* **16**, (2020).
153. Yao, Y. *et al.* A delta-secretase-truncated APP fragment activates CEBPB, mediating Alzheimer’s disease pathologies. *Brain* **144**, 1833–1852 (2021).
154. Rusek, M. *et al.* The Role of the JAK/STAT Signaling Pathway in the Pathogenesis of Alzheimer’s Disease: New Potential Treatment Target. *Int J Mol Sci* **24**, 864 (2023).
155. Ma, Y. *et al.* Epigenetic Regulation of Neuroinflammation in Alzheimer’s Disease. *Cells* **13**, 79 (2023).
156. Xiong, X. *et al.* Epigenomic dissection of Alzheimer’s disease pinpoints causal variants and reveals epigenome erosion. *Cell* **186**, 4422-4437.e21 (2023).

157. Sheehan, B. Assessment scales in dementia. *Ther Adv Neurol Disord* **5**, 349–358 (2012).
158. Morris, J. C. The Clinical Dementia Rating (CDR). *Neurology* **43**, 2412.2–2412-a (1993).
159. Bennett, D. A. *et al.* Religious Orders Study and Rush Memory and Aging Project. *Journal of Alzheimer's Disease* vol. 64 S161–S189 Preprint at <https://doi.org/10.3233/JAD-179939> (2018).
160. La Cognata, V., Morello, G. & Cavallaro, S. Omics data and their integrative analysis to support stratified medicine in neurodegenerative diseases. *International Journal of Molecular Sciences* vol. 22 Preprint at <https://doi.org/10.3390/ijms22094820> (2021).
161. Gerrits, E. *et al.* Distinct amyloid- β and tau-associated microglia profiles in Alzheimer's disease. *Acta Neuropathol* (2021) doi:10.1007/s00401-021-02263-w.
162. Cameron, E. G. *et al.* A molecular switch for neuroprotective astrocyte reactivity. *Nature* **626**, 574–582 (2024).
163. Sierksma, A. *et al.* Novel Alzheimer risk genes determine the microglia response to amyloid- β but not to TAU pathology. *EMBO Mol Med* **12**, (2020).
164. Escartin, C. *et al.* Reactive astrocyte nomenclature, definitions, and future directions. *Nat Neurosci* **24**, 312–325 (2021).
165. Fancy, N. N. *et al.* Characterisation of premature cell senescence in Alzheimer's disease using single nuclear transcriptomics. *Acta Neuropathol* **147**, 78 (2024).
166. Valdés, A. *et al.* Proteomic comparison between different tissue preservation methods for identification of promising biomarkers of urothelial bladder cancer. *Sci Rep* **11**, 7595 (2021).
167. Kong, A. T., Leprevost, F. V., Avtonomov, D. M., Mellacheruvu, D. & Nesvizhskii, A. I. MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat Methods* **14**, 513–520 (2017).
168. Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S. & Ralser, M. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat Methods* **17**, 41–44 (2020).
169. Van den Berge, K. *et al.* Trajectory-based differential expression analysis for single-cell sequencing data. *Nat Commun* **11**, 1201 (2020).
170. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* **2008**, P10008 (2008).
171. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Series B Stat Methodol* **57**, 289–300 (1995).
172. Morabito, S., Reese, F., Rahimzadeh, N., Miyoshi, E. & Swarup, V. hdWGCNA identifies co-expression networks in high-dimensional transcriptomics data. *Cell Reports Methods* **3**, 100498 (2023).
173. Finak, G. *et al.* MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol* **16**, (2015).
174. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLoS Comput Biol* **11**, e1004219 (2015).
175. Jansen, I. E. *et al.* Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat Genet* **51**, 404–413 (2019).

176. Wightman, D. P. *et al.* A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease. *Nat Genet* **53**, 1276–1282 (2021).
177. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics* **81**, 559–575 (2007).
178. Chalise, P. & Fridley, B. L. Integrative clustering of multi-level 'omic data based on non-negative matrix factorization algorithm. *PLoS One* **12**, e0176278 (2017).
179. Meng, C., Helm, D., Frejno, M. & Kuster, B. moCluster: Identifying Joint Patterns Across Multiple Omics Data Sets. *J Proteome Res* **15**, 755–765 (2016).
180. Honer, W. G. *et al.* Cognitive reserve, presynaptic proteins and dementia in the elderly. *Transl Psychiatry* **2**, e114–e114 (2012).
181. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models Using **lme4**. *J Stat Softw* **67**, (2015).
182. Dammer, E. B. *et al.* Multi-platform proteomic analysis of Alzheimer's disease cerebrospinal fluid and plasma reveals network biomarkers associated with proteostasis and the matrisome. *Alzheimers Res Ther* **14**, 174 (2022).
183. Mathys, H. *et al.* Single-cell multiregion dissection of Alzheimer's disease. *Nature* **632**, 858–868 (2024).
184. Johnson, E. C. B. *et al.* Large-scale proteomic analysis of Alzheimer's disease brain and cerebrospinal fluid reveals early changes in energy metabolism associated with microglia and astrocyte activation. *Nat Med* **26**, 769–780 (2020).
185. Caramello, A. *et al.* Intra-cellular accumulation of amyloid is a marker of selective neuronal vulnerability in Alzheimer's disease. doi:10.1101/2023.11.23.23298911.
186. Sturchio, A. *et al.* High cerebrospinal amyloid- β 42 is associated with normal cognition in individuals with brain amyloidosis. *EClinicalMedicine* **38**, 100988 (2021).
187. Kumar, A. *et al.* Genetic effects on longitudinal cognitive decline during the early stages of Alzheimer's disease. *Sci Rep* **11**, 19853 (2021).
188. Yang, H. *et al.* Microglia-specific Alzheimer's disease polygenic risk score is associated with amyloid- β , tau, and microglial activation. *Alzheimer's & Dementia* **18**, (2022).
189. Lin, Y.-T. *et al.* APOE4 Causes Widespread Molecular and Cellular Alterations Associated with Alzheimer's Disease Phenotypes in Human iPSC-Derived Brain Cell Types. *Neuron* **98**, 1141–1154.e7 (2018).
190. Therriault, J. *et al.* Biomarker modeling of Alzheimer's disease using PET-based Braak staging. *Nat Aging* **2**, 526–535 (2022).
191. Leng, K. *et al.* Molecular characterization of selectively vulnerable neurons in Alzheimer's Disease. (2020) doi:10.1101/2020.04.04.025825.
192. Williams, J. B., Cao, Q. & Yan, Z. Transcriptomic analysis of human brains with Alzheimer's disease reveals the altered expression of synaptic genes linked to cognitive deficits. *Brain Commun* **3**, (2021).
193. Jiwaji, Z. & Hardingham, G. E. Good, bad, and neglectful: Astrocyte changes in neurodegenerative disease. *Free Radic Biol Med* **182**, 93–99 (2022).
194. Chen, F. *et al.* Clusterin secreted from astrocyte promotes excitatory synaptic transmission and ameliorates Alzheimer's disease neuropathology. *Mol Neurodegener* **16**, 5 (2021).
195. Harold, D. *et al.* Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat Genet* **41**, 1088–1093 (2009).

196. Zhao, X. *et al.* Alzheimer's disease protective allele of Clusterin modulates neuronal excitability through lipid-droplet-mediated neuron-glia communication. Preprint at <https://doi.org/10.1101/2024.08.14.24312009> (2024).
197. Zhang, H. *et al.* Cerebrospinal fluid synaptosomal-associated protein 25 is a key player in synaptic degeneration in mild cognitive impairment and Alzheimer's disease. *Alzheimers Res Ther* **10**, 80 (2018).
198. Giuffrè, G. M. *et al.* Cerebrospinal fluid neurofilament light chain and total-tau as biomarkers of neurodegeneration in Alzheimer's disease and frontotemporal dementia. *Neurobiol Dis* **186**, 106267 (2023).
199. Qiang, Q., Skudder-Hill, L., Toyota, T., Wei, W. & Adachi, H. CSF GAP-43 as a biomarker of synaptic dysfunction is associated with tau pathology in Alzheimer's disease. *Sci Rep* **12**, 17392 (2022).
200. Hakobyan, S. *et al.* Complement Biomarkers as Predictors of Disease Progression in Alzheimer's Disease. *Journal of Alzheimer's Disease* **54**, 707–716 (2016).
201. Hansson, O. *et al.* CSF biomarkers of Alzheimer's disease concord with amyloid- β PET and predict clinical progression: A study of fully automated immunoassays in BioFINDER and ADNI cohorts. *Alzheimer's & Dementia* **14**, 1470–1481 (2018).
202. Plassman, B. L. *et al.* Incidence of dementia and cognitive impairment, not dementia in the united states. *Ann Neurol* **70**, 418–426 (2011).
203. Dorostkar, M. M., Zou, C., Blazquez-Llorca, L. & Herms, J. Analyzing dendritic spine pathology in Alzheimer's disease: problems and opportunities. *Acta Neuropathol* **130**, 1–19 (2015).
204. Scheff, S. W., DeKosky, S. T. & Price, D. A. Quantitative assessment of cortical synaptic density in Alzheimer's disease. *Neurobiol Aging* **11**, 29–37 (1990).
205. Griffiths, J. & Grant, S. G. N. Synapse pathology in Alzheimer's disease. *Semin Cell Dev Biol* **139**, 13–23 (2023).
206. Stephen, T.-L. *et al.* Imbalance in the response of pre- and post-synaptic components to amyloidopathy. *Sci Rep* **9**, 14837 (2019).
207. Jackson, J. S. *et al.* Altered Synapse Stability in the Early Stages of Tauopathy. *Cell Rep* **18**, 3063–3068 (2017).
208. Jackson, J. *et al.* Targeting the Synapse in Alzheimer's Disease. *Front Neurosci* **13**, (2019).
209. Peng, L., Bestard-Lorigados, I. & Song, W. The synapse as a treatment avenue for Alzheimer's Disease. *Mol Psychiatry* **27**, 2940–2949 (2022).
210. Sorokina, O. *et al.* A unified resource and configurable model of the synapse proteome and its role in disease. *Sci Rep* **11**, 9967 (2021).
211. Gulyácssy, P. *et al.* Proteomic comparison of different synaptosome preparation procedures. *Amino Acids* **52**, 1529–1543 (2020).
212. Allen, N. J. & Eroglu, C. Cell Biology of Astrocyte-Synapse Interactions. *Neuron* **96**, 697–708 (2017).
213. Haydon, P. G. & Carmignoto, G. Astrocyte Control of Synaptic Transmission and Neurovascular Coupling. *Physiol Rev* **86**, 1009–1031 (2006).
214. Imrie, G., Gray, M. B., Raghuraman, V. & Farhy-Tselnicker, I. Gene Expression at the Tripartite Synapse: Bridging the Gap Between Neurons and Astrocytes. in 95–136 (2024). doi:10.1007/978-3-031-64839-7_5.
215. Araque, A., Parpura, V., Sanzgiri, R. P. & Haydon, P. G. Tripartite synapses: glia, the unacknowledged partner. *Trends Neurosci* **22**, 208–215 (1999).

216. Reichenbach, A., Derouiche, A. & Kirchhoff, F. Morphology and dynamics of perisynaptic glia. *Brain Res Rev* **63**, 11–25 (2010).
217. Perez-Alvarez, A., Navarrete, M., Covelo, A., Martin, E. D. & Araque, A. Structural and Functional Plasticity of Astrocyte Processes and Dendritic Spine Interactions. *Journal of Neuroscience* **34**, 12738–12744 (2014).
218. Muñoz-Castro, C. & Serrano-Pozo, A. Astrocyte–Neuron Interactions in Alzheimer’s Disease. in 345–382 (2024). doi:10.1007/978-3-031-64839-7_14.
219. Kumar, A., Fontana, I. C. & Nordberg, A. Reactive astrogliosis: A friend or foe in the pathogenesis of Alzheimer’s disease. *J Neurochem* **164**, 309–324 (2023).
220. Carter, S. F. *et al.* Evidence for Astrocytosis in Prodromal Alzheimer Disease Provided by ¹¹C-Deuterium-L-Deprenyl: A Multitracer PET Paradigm Combining ¹¹C-Pittsburgh Compound B and ¹⁸F-FDG. *Journal of Nuclear Medicine* **53**, 37–46 (2012).
221. Nam, M.-H. *et al.* Visualizing reactive astrocyte-neuron interaction in Alzheimer’s disease using ¹¹C-acetate and ¹⁸F-FDG. *Brain* **146**, 2957–2974 (2023).
222. Sadick, J. S. *et al.* Astrocytes and oligodendrocytes undergo subtype-specific transcriptional changes in Alzheimer’s disease. *Neuron* **110**, 1788–1805.e10 (2022).
223. Labib, D. *et al.* Proteomic Alterations and Novel Markers of Neurotoxic Reactive Astrocytes in Human Induced Pluripotent Stem Cell Models. *Front Mol Neurosci* **15**, (2022).
224. Clayton, B. L. L. *et al.* A phenotypic screening platform for identifying chemical modulators of astrocyte reactivity. *Nat Neurosci* **27**, 656–665 (2024).
225. Grubman, A. *et al.* A single-cell atlas of entorhinal cortex from individuals with Alzheimer’s disease reveals cell-type-specific gene expression regulation. *Nat Neurosci* **22**, 2087–2097 (2019).
226. Leng, K. *et al.* CRISPRi screens in human iPSC-derived astrocytes elucidate regulators of distinct inflammatory reactive states. *Nat Neurosci* **25**, 1528–1542 (2022).
227. Tzioras, M., McGeachan, R. I., Durrant, C. S. & Spires-Jones, T. L. Synaptic degeneration in Alzheimer disease. *Nat Rev Neurol* **19**, 19–38 (2023).
228. Carter, S. F. *et al.* Astrocyte Biomarkers in Alzheimer’s Disease. *Trends Mol Med* **25**, 77–95 (2019).
229. Mookherjee, P. *et al.* GLT-1 Loss Accelerates Cognitive Deficit Onset in an Alzheimer’s Disease Animal Model. *Journal of Alzheimer’s Disease* **26**, 447–455 (2011).
230. Kobayashi, E. *et al.* Activated forms of astrocytes with higher GLT-1 expression are associated with cognitive normal subjects with Alzheimer pathology in human brain. *Sci Rep* **8**, 1712 (2018).
231. Nanclares, C., Baraibar, A. M., Araque, A. & Kofuji, P. Dysregulation of Astrocyte–Neuronal Communication in Alzheimer’s Disease. *Int J Mol Sci* **22**, 7887 (2021).
232. Sompol, P. *et al.* Calcineurin/NFAT Signaling in Activated Astrocytes Drives Network Hyperexcitability in Aβ-Bearing Mice. *The Journal of Neuroscience* **37**, 6132–6148 (2017).
233. Yi, C. *et al.* Astroglial connexin43 contributes to neuronal suffering in a mouse model of Alzheimer’s disease. *Cell Death Differ* **23**, 1691–1701 (2016).
234. Balu, D. T. *et al.* Neurotoxic astrocytes express the d-serine synthesizing enzyme, serine racemase, in Alzheimer’s disease. *Neurobiol Dis* **130**, 104511 (2019).

235. Pirttimäki, T. M. *et al.* $\alpha 7$ Nicotinic Receptor-Mediated Astrocytic Gliotransmitter Release: A β Effects in a Preclinical Alzheimer's Mouse Model. *PLoS One* **8**, e81828 (2013).
236. Yoon, B. *et al.* Glial GABA, synthesized by monoamine oxidase B, mediates tonic inhibition. *J Physiol* **592**, 4951–4968 (2014).
237. Jaisa-aad, M., Muñoz-Castro, C., Healey, M. A., Hyman, B. T. & Serrano-Pozo, A. Characterization of monoamine oxidase-B (MAO-B) as a biomarker of reactive astrogliosis in Alzheimer's disease and related dementias. *Acta Neuropathol* **147**, 66 (2024).
238. Jo, S. *et al.* GABA from reactive astrocytes impairs memory in mouse models of Alzheimer's disease. *Nat Med* **20**, 886–896 (2014).
239. Park, J.-H. *et al.* Newly developed reversible MAO-B inhibitor circumvents the shortcomings of irreversible inhibitors in Alzheimer's disease. *Sci Adv* **5**, (2019).
240. Chun, H. *et al.* Severe reactive astrocytes precipitate pathological hallmarks of Alzheimer's disease via H₂O₂– production. *Nat Neurosci* **23**, 1555–1566 (2020).
241. Bi, F. *et al.* Reactive astrocytes secrete lcn2 to promote neuron death. *Proceedings of the National Academy of Sciences* **110**, 4069–4074 (2013).
242. Murray, T. E. *et al.* Potential neurotoxic activity of diverse molecules released by astrocytes. *Brain Res Bull* **189**, 80–101 (2022).
243. Zeppenfeld, D. M. *et al.* Association of Perivascular Localization of Aquaporin-4 With Cognition and Alzheimer Disease in Aging Brains. *JAMA Neurol* **74**, 91 (2017).
244. Smith, A. D. Imaging the progression of Alzheimer pathology through the brain. *Proceedings of the National Academy of Sciences* **99**, 4135–4137 (2002).
245. Carlyle, B. C. *et al.* Synaptic proteins associated with cognitive performance and neuropathology in older humans revealed by multiplexed fractionated proteomics. *Neurobiol Aging* **105**, 99–114 (2021).
246. Koopmans, F. *et al.* SynGO: An Evidence-Based, Expert-Curated Knowledge Base for the Synapse. *Neuron* **103**, 217–234.e4 (2019).
247. Galea, E. *et al.* Multi-transcriptomic analysis points to early organelle dysfunction in human astrocytes in Alzheimer's disease. *Neurobiol Dis* **166**, 105655 (2022).
248. Niu, M. *et al.* Droplet-based transcriptome profiling of individual synapses. *Nat Biotechnol* **41**, 1332–1344 (2023).
249. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-Seq data. *BMC Bioinformatics* **14**, 7 (2013).
250. Browaeys, R., Saelens, W. & Saeys, Y. NicheNet: modeling intercellular communication by linking ligands to target genes. *Nat Methods* **17**, 159–162 (2020).
251. Laszlo, Z. I. *et al.* Synaptic proteomics reveal distinct molecular signatures of cognitive change and C9ORF72 repeat expansion in the human ALS cortex. *Acta Neuropathol Commun* **10**, 156 (2022).
252. Swarbrick, S., Wragg, N., Ghosh, S. & Stolzing, A. Systematic Review of miRNA as Biomarkers in Alzheimer's Disease. *Mol Neurobiol* **56**, 6156–6167 (2019).
253. Cai, Q. & Tammineni, P. Mitochondrial Aspects of Synaptic Dysfunction in Alzheimer's Disease. *J Alzheimers Dis* **57**, 1087–1103 (2017).
254. Kurucu, H. *et al.* Inhibitory synapse loss and accumulation of amyloid beta in inhibitory presynaptic terminals in Alzheimer's disease. *Eur J Neurol* (2021) doi:10.1111/ene.15043.

255. Ross, S. P. *et al.* miRNA-431 Prevents Amyloid- β -Induced Synapse Loss in Neuronal Cell Culture Model of Alzheimer's Disease by Silencing Kremen1. *Front Cell Neurosci* **12**, 87 (2018).
256. Martin Flores, N. *et al.* Downregulation of Dickkopf-3, a Wnt antagonist elevated in Alzheimer's disease, restores synapse integrity and memory in a disease mouse model. *Elife* **12**, (2024).
257. Reddy, P. H. *et al.* Abnormal mitochondrial dynamics and synaptic degeneration as early events in Alzheimer's disease: implications to mitochondria-targeted antioxidant therapeutics. *Biochim Biophys Acta* **1822**, 639–649 (2012).
258. Myeong, J., Stunault, M. I., Klyachko, V. A. & Ashrafi, G. Metabolic regulation of single synaptic vesicle exo- and endocytosis in hippocampal synapses. *Cell Rep* **43**, 114218 (2024).
259. Andreassen, M. & Nedergaard, S. Effect of acute mitochondrial dysfunction on hyperexcitable network activity in rat hippocampus in vitro. *Brain Res* **1751**, 147193 (2021).
260. Mendonça, C. F. *et al.* Proteomic signatures of brain regions affected by tau pathology in early and late stages of Alzheimer's disease. *Neurobiol Dis* **130**, 104509 (2019).
261. Sanchez-Rodriguez, L. M. *et al.* Revealing the combined roles of A β and tau in Alzheimer's disease via a pathophysiological activity decoder. *bioRxiv* (2023) doi:10.1101/2023.02.21.529377.
262. Scaduto, P. *et al.* Functional excitatory to inhibitory synaptic imbalance and loss of cognitive performance in people with Alzheimer's disease neuropathologic change. *Acta Neuropathol* **145**, 303–324 (2023).
263. Lavialle, M. *et al.* Structural plasticity of perisynaptic astrocyte processes involves ezrin and metabotropic glutamate receptors. *Proceedings of the National Academy of Sciences* **108**, 12915–12919 (2011).
264. Michaluk, P., Heller, J. P. & Rusakov, D. A. Rapid recycling of glutamate transporters on the astroglial surface. *Elife* **10**, (2021).
265. Rapino, F. *et al.* Small-molecule screen reveals pathways that regulate C4 secretion in stem cell-derived astrocytes. *Stem Cell Reports* **18**, 237–253 (2023).
266. Mou, T.-C. M. *et al.* Association of complement component 4 with neuroimmune abnormalities in the subventricular zone in schizophrenia and autism spectrum disorders. *Neurobiol Dis* **173**, 105840 (2022).
267. Luo, J. Dysregulation of polarity proteins in astrocyte reactivity. *Ageing Res Rev* **86**, 101869 (2023).
268. Sakers, K. *et al.* Astrocytes locally translate transcripts in their peripheral processes. *Proceedings of the National Academy of Sciences* **114**, (2017).
269. Murata, K. *et al.* The extracellular fragment of GPNMB (Glycoprotein nonmelanosoma protein B, osteoactivin) improves memory and increases hippocampal GluA1 levels in mice. *J Neurochem* **132**, 583–594 (2015).
270. Negro, D. & Opazo, P. Cognitive resilience in Alzheimer's disease: from large-scale brain networks to synapses. *Brain Commun* **6**, (2023).
271. Bhembre, N., Bonthron, C. & Opazo, P. Synaptic Compensatory Plasticity in Alzheimer's Disease. *The Journal of Neuroscience* **43**, 6833 (2023).
272. Hruska, M., Cain, R. E. & Dalva, M. B. Nanoscale rules governing the organization of glutamate receptors in spine synapses are subunit specific. *Nat Commun* **13**, 920 (2022).

273. DiGregorio, D. A., Nusser, Z. & Silver, R. A. Spillover of Glutamate onto Synaptic AMPA Receptors Enhances Fast Transmission at a Cerebellar Synapse. *Neuron* **35**, 521–533 (2002).
274. Henneberger, C. *et al.* LTP Induction Boosts Glutamate Spillover by Driving Withdrawal of Perisynaptic Astroglia. *Neuron* **108**, 919–936.e11 (2020).
275. Stafford, M. M., Brown, M. N., Mishra, P., Stanwood, G. D. & Mathews, G. C. Glutamate spillover augments GABA synthesis and release from axodendritic synapses in rat hippocampus. *Hippocampus* **20**, 134–144 (2010).
276. Hobson, B. D. & Herzog, E. Methodological concerns and lack of evidence for single-synapse RNA-seq. *Nat Biotechnol* **41**, 1221–1224 (2023).
277. Jackson, R. J. *et al.* Clusterin accumulates in synapses in Alzheimer’s disease and is increased in apolipoprotein E4 carriers. *Brain Commun* **1**, (2019).
278. Morrison, J. H. & Baxter, M. G. The ageing cortical synapse: hallmarks and implications for cognitive decline. *Nat Rev Neurosci* **13**, 240–250 (2012).
279. Ali, H. R. *et al.* Imaging mass cytometry and multiplatform genomics define the phenogenomic landscape of breast cancer. *Nat Cancer* **1**, 163–175 (2020).
280. Elaldi, R. *et al.* High Dimensional Imaging Mass Cytometry Panel to Visualize the Tumor Immune Microenvironment Contexture. *Front Immunol* **12**, (2021).
281. Marshall, C. R., Farrow, M. A., Djambazova, K. V. & Spraggins, J. M. Untangling Alzheimer’s disease with spatial multi-omics: a brief review. *Frontiers in Aging Neuroscience* vol. 15 Preprint at <https://doi.org/10.3389/fnagi.2023.1150512> (2023).
282. Bressan, D., Battistoni, G. & Hannon, G. J. The dawn of spatial omics. *Science* (1979) **381**, (2023).
283. Ramaglia, V. *et al.* Multiplexed imaging of immune cells in staged multiple sclerosis lesions by mass cytometry. *Elife* **8**, (2019).
284. Chen, W.-T. *et al.* Spatial Transcriptomics and In Situ Sequencing to Study Alzheimer’s Disease. *Cell* **182**, 976–991.e19 (2020).
285. Choi, H. *et al.* Spatiotemporal characterization of glial cell activation in an Alzheimer’s disease model by spatially resolved transcriptomics. *Exp Mol Med* **55**, 2564–2575 (2023).
286. Johnston, K. G. *et al.* Single-cell spatial transcriptomics reveals distinct patterns of dysregulation in non-neuronal and neuronal cells induced by the Trem2R47H Alzheimer’s risk gene mutation. *Mol Psychiatry* (2024) doi:10.1038/s41380-024-02651-0.
287. Chen, S. *et al.* Spatially resolved transcriptomics reveals genes associated with the vulnerability of middle temporal gyrus in Alzheimer’s disease. *Acta Neuropathol Commun* **10**, 188 (2022).
288. Avey, D. R. *et al.* Uncovering Plaque-Glia Niches in Human Alzheimer’s Disease Brains Using Spatial Transcriptomics. *bioRxiv* 2024.09.05.611566 (2024) doi:10.1101/2024.09.05.611566.
289. Jury-Garfe, N. *et al.* Enhanced microglial dynamics and a paucity of tau seeding in the amyloid plaque microenvironment contribute to cognitive resilience in Alzheimer’s disease. *Acta Neuropathol* **148**, 15 (2024).
290. Muñoz-Castro, C. *et al.* Cyclic multiplex fluorescent immunohistochemistry and machine learning reveal distinct states of astrocytes and microglia in normal aging and Alzheimer’s disease. *J Neuroinflammation* **19**, 30 (2022).

291. Thal, D. R., Capetillo-Zarate, E., Del Tredici, K. & Braak, H. The Development of Amyloid β Protein Deposits in the Aged Brain. *Science of Aging Knowledge Environment* **2006**, (2006).
292. DeTure, M. A. & Dickson, D. W. The neuropathological diagnosis of Alzheimer's disease. *Mol Neurodegener* **14**, 32 (2019).
293. Baghallab, I., Reyes-Ruiz, J. M., Abulnaja, K., Huwait, E. & Glabe, C. Epitomic Characterization of the Specificity of the Anti-Amyloid A β Monoclonal Antibodies 6E10 and 4G8. *Journal of Alzheimer's Disease* **66**, 1235–1244 (2018).
294. Malek-Ahmadi, M., Perez, S. E., Chen, K. & Mufson, E. J. Neuritic and Diffuse Plaque Associations with Memory in Non-Cognitively Impaired Elderly. *Journal of Alzheimer's Disease* **53**, 1641–1652 (2016).
295. Goedert, M., Jakes, R. & Vanmechelen, E. Monoclonal antibody AT8 recognises tau protein phosphorylated at both serine 202 and threonine 205. *Neurosci Lett* **189**, 167–170 (1995).
296. Thal, D. R. *et al.* Sequence of A β -Protein Deposition in the Human Medial Temporal Lobe. *J Neuropathol Exp Neurol* **59**, 733–748 (2000).
297. Schubert, P., Ogata, T., Marchini, C. & Ferroni, S. Glia-related pathomechanisms in Alzheimer's disease: a therapeutic target? *Mech Ageing Dev* **123**, 47–57 (2001).
298. McGEER, P. L., ITAGAKI, S., TAGO, H. & McGEER, E. G. Occurrence of HLA-DR Reactive Microglia in Alzheimer's Disease. *Ann N Y Acad Sci* **540**, 319–323 (1988).
299. Heneka, M. T. *et al.* Neuroinflammation in Alzheimer's disease. *The Lancet Neurology* vol. 14 388–405 Preprint at [https://doi.org/10.1016/S1474-4422\(15\)70016-5](https://doi.org/10.1016/S1474-4422(15)70016-5) (2015).
300. Windhager, J., Bodenmiller, B. & Eling, N. An end-to-end workflow for multiplexed image processing and analysis. doi:10.1101/2021.11.12.468357.
301. Greenwald, N. F. *et al.* Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. *Nat Biotechnol* **40**, 555–565 (2022).
302. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676–682 (2012).
303. Windhager, J. *et al.* An end-to-end workflow for multiplexed image processing and analysis. *Nat Protoc* **18**, 3565–3613 (2023).
304. Righelli, D. *et al.* SpatialExperiment: infrastructure for spatially-resolved transcriptomics data in R using Bioconductor. *Bioinformatics* **38**, 3128–3131 (2022).
305. Chevrier, S. *et al.* Compensation of Signal Spillover in Suspension and Imaging Mass Cytometry. *Cell Syst* **6**, 612–620.e5 (2018).
306. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **36**, 411–420 (2018).
307. Miller, S. A. & Policastro, R. A. scProportionTest. Preprint at <https://doi.org/https://github.com/rpolicastro/scProportionTest> (2020).
308. Becht, E. *et al.* Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol* **37**, 38–44 (2019).
309. Eling, N., Damond, N., Hoch, T. & Bodenmiller, B. cytomapper: An R/Bioconductor package for visualization of highly multiplexed imaging data. *Bioinformatics* **36**, 5706–5708 (2020).

310. Schapiro, D. *et al.* histoCAT: analysis of cell phenotypes and interactions in multiplex image cytometry data. *Nat Methods* **14**, 873–876 (2017).
311. Zhang, Z. *et al.* The Appropriate Marker for Astrocytes: Comparing the Distribution and Expression of Three Astrocytic Markers in Different Mouse Cerebral Regions. *Biomed Res Int* **2019**, 1–15 (2019).
312. Hannon, E. *et al.* Quantifying the proportion of different cell types in the human cortex using DNA methylation profiles. *BMC Biol* **22**, 17 (2024).
313. Habib, N. *et al.* Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat Methods* **14**, 955–958 (2017).
314. Ding, J. *et al.* Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. *Nat Biotechnol* **38**, 737–746 (2020).
315. Zhou, Y. *et al.* Human and mouse single-nucleus transcriptomics reveal TREM2-dependent and TREM2-independent cellular responses in Alzheimer’s disease. *Nat Med* **26**, 131–142 (2020).
316. Wang, Y. *et al.* TREM2-mediated early microglial response limits diffusion and toxicity of amyloid plaques. *Journal of Experimental Medicine* **213**, 667–675 (2016).
317. Leyns, C. E. G. *et al.* TREM2 function impedes tau seeding in neuritic plaques. *Nat Neurosci* **22**, 1217–1222 (2019).
318. Wang, C. *et al.* Astrocyte-TREM2 alleviates brain injury by regulating reactive astrocyte states following ischemic stroke. *Glia* **72**, 2061–2078 (2024).
319. Milosevic, V. Different approaches to Imaging Mass Cytometry data analysis. *Bioinformatics Advances* vol. 3 Preprint at <https://doi.org/10.1093/bioadv/vbad046> (2023).
320. Hong, S. *et al.* Complement and microglia mediate early synapse loss in Alzheimer mouse models. *Science (1979)* **352**, 712–716 (2016).
321. Shi, Y. & Holtzman, D. M. Interplay between innate immunity and Alzheimer disease: APOE and TREM2 in the spotlight. *Nat Rev Immunol* **18**, 759–772 (2018).
322. Mecca, A. P. *et al.* Synaptic density and cognitive performance in Alzheimer’s disease: A PET imaging study with [11 C]UCB-J. *Alzheimers Dement* **18**, 2527–2536 (2022).
323. Thrupp, N. *et al.* Single-Nucleus RNA-Seq Is Not Suitable for Detection of Microglial Activation Genes in Humans. *Cell Rep* **32**, (2020).
324. Bakken, T. E. *et al.* Single-nucleus and single-cell transcriptomes compared in matched cortical cell types. *PLoS One* **13**, e0209648 (2018).
325. Vandereyken, K., Sifrim, A., Thienpont, B. & Voet, T. Methods and applications for single-cell and spatial multi-omics. *Nat Rev Genet* **24**, 494–515 (2023).
326. Neff, R. A. *et al.* Molecular Subtyping of Alzheimer’s Disease Using RNA Sequencing Data Reveals Novel Mechanisms and Targets. *Sci. Adv* vol. 7 <https://www.science.org> (2021).
327. Haglund, A. *et al.* Cell state-dependent allelic effects and contextual Mendelian randomization analysis for human brain phenotypes. *Nat Genet* (2025) doi:10.1038/s41588-024-02050-9.
328. Lee, L. Y. *et al.* Robust and interpretable AI-guided marker for early dementia prediction in real-world clinical settings. *EClinicalMedicine* **74**, 102725 (2024).
329. Gaitsch, H., Franklin, R. J. M. & Reich, D. S. Cell-free DNA-based liquid biopsies in neurology. *Brain* **146**, 1758–1774 (2023).
330. Baligács, N. *et al.* Homeostatic microglia initially seed and activated microglia later reshape amyloid plaques in Alzheimer’s Disease. *Nat Commun* **15**, 10634 (2024).

- 331. Paul, R. *et al.* INVOKE-2: A phase 2 randomized, double-blind, placebo-controlled study to evaluate the efficacy and safety of AL002 in participants with early Alzheimer's disease. *Alzheimer's & Dementia* **17**, (2021).
- 332. Yi-Bin, W. *et al.* Inhibition of the CEBP β -NF κ B interaction by nanocarrier-packaged Carnosic acid ameliorates glia-mediated neuroinflammation and improves cognitive function in an Alzheimer's disease model. *Cell Death Dis* **13**, 318 (2022).
- 333. Mehla, J. *et al.* STAT3 inhibitor mitigates cerebral amyloid angiopathy and parenchymal amyloid plaques while improving cognitive functions and brain networks. *Acta Neuropathol Commun* **9**, 193 (2021).
- 334. Furman, J. L. *et al.* Targeting Astrocytes Ameliorates Neurologic Changes in a Mouse Model of Alzheimer's Disease. *The Journal of Neuroscience* **32**, 16129–16140 (2012).
- 335. Kim, S. *et al.* Astrocytic autophagy plasticity modulates A β clearance and cognitive function in Alzheimer's disease. *Mol Neurodegener* **19**, 55 (2024).